

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-97241

(43) 公開日 平成9年(1997)4月8日

(51) Int.Cl.<sup>6</sup>

G 0 6 F 15/18

識別記号

庁内整理番号

F I

G 0 6 F 15/18

4 2 0 C

E

1/28

1/00

3 3 4 H

技術表示箇所

審査請求 未請求 請求項の数10 O L (全 34 頁)

(21) 出願番号

特願平7-251427

(22) 出願日

平成7年(1995)9月28日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 松下 祐之

神奈川県海老名市下今泉810番地 株式会

社日立製作所オフィスシステム事業部内

(72) 発明者 宇賀神 教

神奈川県海老名市下今泉810番地 株式会

社日立製作所オフィスシステム事業部内

(74) 代理人 弁理士 秋田 収喜

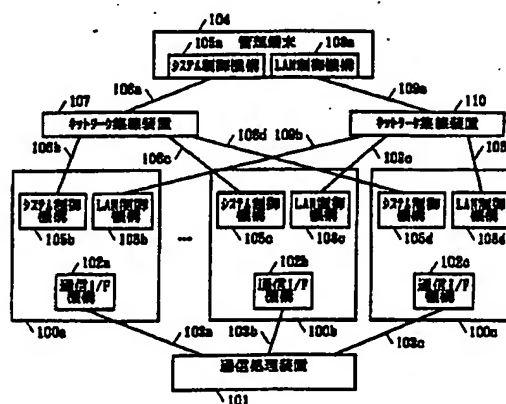
(54) 【発明の名称】 並列計算機システムの管理装置

(57) 【要約】

【課題】 並列計算機システムを構成する複数のノードの運用管理を管理端末装置で一括して行う。

【解決手段】 複数のノードに、各ノードの主電源により動作し並列処理を実行するメインプロセッサと、各ノードの補助電源により動作し前記メインプロセッサを管理するシステム制御コマンドを実行するサブプロセッサと、前記補助電源により動作し管理端末装置との通信を行うシステム制御機構とを備え、管理端末装置に、前記複数のノードの複数のシステム制御機構と通信を行うシステム制御機構を備え、前記複数のノードの複数のシステム制御機構と前記管理端末装置のシステム制御機構とを接続して成り、前記メインプロセッサを管理するシステム制御コマンドを前記管理端末装置から前記サブプロセッサに送信するシステム制御インタフェースを備えるものである。

図1



## 【特許請求の範囲】

【請求項1】 複数の計算機であるノードを接続した並列計算機システムを管理する管理端末装置を備えた並列計算機システムの管理装置において、

前記複数のノードは、当該ノードの主電源により動作し並列処理を実行するメインプロセッサと、当該ノードの補助電源により動作し前記メインプロセッサを管理するシステム制御コマンドを実行するサブプロセッサと、前記補助電源により動作し前記管理端末装置との通信を行うシステム制御機構とを備え、

前記管理端末装置は、前記複数のノードの複数のシステム制御機構と通信を行うシステム制御機構を備え、

前記複数のノードの複数のシステム制御機構と前記管理端末装置のシステム制御機構とを接続して成り、前記複数のノードの複数のメインプロセッサを管理するシステム制御コマンドを前記管理端末装置から前記複数のノードの複数のサブプロセッサに送信するシステム制御インタフェースを備えることを特徴とする並列計算機システムの管理装置。

【請求項2】 前記複数のノードのサブプロセッサは、当該ノードの主電源を投入または切断する機能を備え、前記管理端末装置は、前記複数のノードのサブプロセッサに、一括または個別に主電源を投入または切断するシステム制御コマンドを送信する手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

【請求項3】 前記管理端末装置は、前記複数のノードの主電源を個別に投入するシステム制御コマンドを、予め設定された時間間隔で前記複数のノードのサブプロセッサに個別に送信する手段を備えることを特徴とする請求項2に記載された並列計算機システムの管理装置。

【請求項4】 前記管理端末装置は、前記複数のノードの特定のノードのサブプロセッサに特定のシステム制御コマンドを送信し、予め設定された時間内に前記特定のシステム制御コマンドに対する正常な応答が受信されない場合に、前記特定のノードに異常が発生しているとみなす手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

【請求項5】 前記複数のノードのシステム制御機構は、当該ノードのメインプロセッサまたはサブプロセッサが動作時に出力するメッセージであるノードメッセージを蓄積する手段を備え、前記管理端末装置は、当該ノードのシステム制御機構に蓄積されたノードメッセージを読み取る手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

【請求項6】 前記複数のノードのサブプロセッサは、当該ノードのメインメモリまたはレジスタの内容を参照及び更新する手段を備え、前記管理端末装置は、前記複数のノードのサブプロセッサに、当該ノードのメインメモリまたはレジスタの内容を参照または更新するシステ

ム制御コマンドを送信する手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

【請求項7】 前記複数のノードのサブプロセッサは、当該ノードのメインプロセッサをリセットする手段を備え、前記管理端末装置は、当該ノードのサブプロセッサに、当該ノードのメインプロセッサをリセットするシステム制御コマンドを送信する手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

【請求項8】 前記複数のノードのサブプロセッサは、当該ノードのメインメモリの内容を参照及び更新する手段と、当該ノードのメインプロセッサをリセットする手段とを備え、前記管理端末装置は、当該ノードのメインプロセッサが格納しているメインメモリ中のブートストラップデバイス名を参照及び更新するシステム制御コマンドと、当該ノードのメインプロセッサをリセットするシステム制御コマンドとを送信する手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

【請求項9】 前記管理端末装置を複数備え、前記複数の管理端末装置のうちの一部の管理端末装置の機能を制限する手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

【請求項10】 前記管理端末装置は、補助電源で動作し、特定の信号を入力すると前記管理端末装置の主電源を投入する電源投入論理と、前記電源投入論理により主電源が投入されたときに、前記複数のノードのサブプロセッサに、一括または個別に主電源を投入するシステム制御コマンドを送信する手段を備えることを特徴とする請求項1に記載された並列計算機システムの管理装置。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、並列計算機システムの管理装置に関し、特に、並列計算機システムを構成する複数のノードのメインプロセッサが動作していない場合であっても前記複数のノードの保守及び管理を行う並列計算機システムの管理装置に適用して有効な技術に関するものである。

【0002】

【従来の技術】従来、複数の計算機で構成される計算機システムの運用管理方法及びその実施装置について、いくつかのものが提案されている。

【0003】複数のUNIXマシンのコンソールを1台にまとめたときに発生する運用と監視の負荷増大を防止する複数のUNIXマシンの集中運用および監視コンソールディスプレイについては、特開平6-214763号公報に記載されている。

【0004】その概要は、複数のUNIXマシンを集中運用及び管理するサーバーであるセンター・コンソール

10

20

30

40

50

に、運用目的別にコマンドの宛先を格納した宛先テーブルを作成しておき、前記宛先テーブルに従ってコマンドを実行するものである。

【0005】複数の計算機から構成される複合計算機システムにおいて、単一のシステムコンソールにより接続する計算機を切り換えて保守及び操作を行った場合の誤操作を防止する複合計算機システムにおけるコンソール切替制御方式については、特開平5-120247号公報に記載されている。

【0006】その概要は、複数の計算機内のサービスプロセッサ同士を切替装置に接続し、前記切替装置にシステムコンソールを接続し、前記システムコンソールからは、計算機を識別する識別子を用いて、メッセージ出力対象の計算機を順次切り替えていくことにより、複数の計算機で1台のシステムコンソールを共有する方式であり、システムコンソールにより保守及び操作を行う際に、操作を行おうとしている計算機の識別子と、システムコンソールに接続されている計算機の識別子と比較し、識別子が一致する場合に操作を実行するものである。

【0007】分散処理システムを構成する各計算機からのメッセージを集中管理するメッセージ集中管理方式については、特開平5-20281号公報に記載されている。

【0008】その概要は、ネットワークにて接続された複数の計算機内で集中管理ノードを決定し、その集中管理ノードが監視対象ノードの発行する稼働状況メッセージを集中管理する方式である。

【0009】

【発明が解決しようとする課題】本発明者は、前記従来技術を検討した結果、以下の問題点を見出した。

【0010】すなわち、前記従来の複数の計算機から成る計算機システムの管理装置では、管理対象の計算機上で動作しているネットワークソフトウェアの機能を使用して、管理対象の計算機が動作していない場合やオペレーティングシステムが動作していない場合及びネットワークソフトウェアが動作していない場合には、運用管理を行えないという問題があった。

【0011】前記従来の複数のUNIXマシンの集中運用および監視コンソールディスプレイを使用する方法では、管理対象となる計算機は、オペレーティングシステムのUNIXが動作していることが前提となる為、オペレーティングシステムが動作していない場合には、コンソールディスプレイから集中運用および監視ができないという問題があった。

【0012】前記従来の複合計算機システムにおけるコンソール切替制御方式では、システムコンソールと各々のサービスプロセッサとの間に切替装置が存在している為、切替装置なる特別なハードウェアが必要となるという問題があった。

【0013】前記従来のメッセージ集中管理方式では、複数の計算機から集中管理ノードにメッセージが送られてくる為、前記集中管理ノードがシステムダウンとなったときには、メッセージの集中管理が行えないという問題と、前記メッセージは、ノードが接続されるネットワーク経由で送信されてくる為、各ノードのオペレーティングシステム及びネットワークが起動されていない場合、集中管理ノードから各ノードの状態を管理することが出来ないという問題があった。

【0014】本発明の目的は、並列処理を実行するメインプロセッサの動作並びに前記メインプロセッサのオペレーティングシステム及びネットワークソフトウェアの動作とは無関係に、並列計算機システムを構成する複数のノードの運用管理を管理端末装置で一括して行うことが可能な技術を提供することにある。

【0015】本発明の他の目的は、並列計算機システムを構成する複数のノードの電源の投入または切断を管理端末装置で一括または個別に行うことが可能な技術を提供することにある。

【0016】本発明の他の目的は、並列計算機システムに電力を供給する電源設備の突入電流を低く抑えることが可能な技術を提供することにある。

【0017】本発明の他の目的は、並列計算機システムを構成する複数のノードが正常に動作中であることを管理端末装置で監視することが可能な技術を提供することにある。

【0018】本発明の他の目的は、並列計算機システムを構成する複数のノードのメインプロセッサがノードメッセージを出力した後にその動作を停止した場合であっても、前記ノードメッセージを管理端末装置で一括して管理することが可能な技術を提供することにある。

【0019】本発明の他の目的は、並列計算機システムを構成する複数のノードの障害発生時のメインメモリ及びレジスタの内容を管理端末装置で一括して管理することが可能な技術を提供することにある。

【0020】本発明の他の目的は、並列計算機システムを構成する複数のノードのメインプロセッサのリセットを管理端末装置から一括して行うことが可能な技術を提供することにある。

【0021】本発明の他の目的は、並列計算機システムを構成する複数のノードの特定のブートストラップデバイスに障害が発生した場合に、管理端末装置からの指示により、ブートストラップデバイスを変更して前記複数のノードのメインプロセッサのシステム立ち上げ処理を行うことが可能な技術を提供することにある。

【0022】本発明の他の目的は、並列計算機システムを複数の管理端末装置で管理した場合に、前記複数の管理端末装置の動作の競合を防止することが可能な技術を提供することにある。

【0023】本発明の他の目的は、並列計算機システム

の運用管理を遠隔地から行うことが可能な技術を提供することにある。

【0024】本発明の前記並びにその他の目的と新規な特徴は、本明細書の記述及び添付図面によって明らかになるであろう。

【0025】

【課題を解決するための手段】本願において開示される発明のうち、代表的なものの概要を簡単に説明すれば、下記のとおりである。

【0026】(1) 複数の計算機であるノードを接続した並列計算機システムを管理する管理端末装置を備えた並列計算機システムの管理装置において、前記複数のノードは、各ノードの主電源により動作し並列処理を実行するメインプロセッサと、各ノードの補助電源により動作し前記メインプロセッサを管理するシステム制御コマンドを実行するサブプロセッサと、前記補助電源により動作し前記管理端末装置との通信を行うシステム制御機構とを備え、前記管理端末装置は、前記複数のノードの複数のシステム制御機構と通信を行うシステム制御機構を備え、前記複数のノードの複数のシステム制御機構と前記管理端末装置のシステム制御機構とを接続して成り、前記複数のノードの複数のメインプロセッサを管理するシステム制御コマンドを前記管理端末装置から前記複数のノードの複数のサブプロセッサに送信するシステム制御インタフェースを備えるものである。

【0027】前記(1)の並列計算機システムの管理装置では、管理端末装置から発行されたシステム制御コマンドは、システム制御インタフェースを構成する前記管理端末装置及び前記複数のノードのシステム制御機構を介し、前記複数のノードのサブプロセッサに送られ、前記サブプロセッサに送られたシステム制御コマンドを、前記サブプロセッサで実行することにより、前記メインプロセッサの運用管理を行う。

【0028】従来の並列計算機システムの管理装置では、前記並列計算機システムの通常業務である並列処理を実行するメインプロセッサで動作している汎用のオペレーティングシステムや、そのオペレーティングシステムの管理下で動作するネットワークソフトウェアを使用して、並列計算機システムを構成する複数のノードの運用管理を行っている。

【0029】この為、前記従来の並列計算機システムの管理装置を使用する場合には、管理対象である並列計算機システムを構成する複数のノードのメインプロセッサが正常に動作し、前記の汎用のオペレーティングシステムやネットワークソフトウェアが実行中であることが前提条件となり、前記管理対象の複数のノードのメインプロセッサが動作していない場合や前記の汎用のオペレーティングシステムやネットワークソフトウェアが動作していない場合、例えば、並列計算機システムの電源が投入されていない運用開始前の状態、または、オペレーテ

ィングシステムやネットワークの構成を変更し、正常に動作するかどうか確かめようとしている状態、或いは、障害の発生により動作しなくなった特定のノードの状況を調査する場合などでは、前記従来の並列計算機システムの管理装置を使用することができなかった。

【0030】そこで、前記(1)の並列計算機システムの管理装置では、メインプロセッサの動作状況とは無関係に動作可能なサブプロセッサとシステム制御機構を、前記(1)の並列計算機システムを構成する全てのノードに備え、各々のノードのシステム制御機構をネットワーク集線装置等の装置を介し、管理端末装置のシステム制御機構に接続している。

【0031】前記複数のノードの各ノードに備えられたサブプロセッサ及びシステム制御機構は、メインプロセッサが使用する主電源とは別の補助電源により動作し、また、前記システム制御機構は、メインプロセッサで動作するネットワークソフトウェア及びそのネットワークソフトウェアが使用する通信ケーブルとは別のネットワークソフトウェア及び通信ケーブルを使用して管理端末装置と通信を行う。

【0032】従って、メインプロセッサが動作していなくても、補助電源によりサブプロセッサ及びシステム制御機構が動作していれば、メインプロセッサの制御を前記管理端末装置から行うことが可能である。

【0033】以上の様に、前記並列計算機システムの管理装置によれば、各ノードの補助電源で動作し、前記メインプロセッサが使用するネットワークソフトウェア及び通信ケーブルとは別のネットワークソフトウェア及び通信ケーブルを使用して管理端末装置と通信を行うシステム制御機構に、前記管理端末装置からシステム制御コマンドを送信し、前記システム制御コマンドを前記補助電源で動作するサブプロセッサで実行することにより複数のノードのメインプロセッサの制御を行うので、並列処理を実行するメインプロセッサの動作並びに前記メインプロセッサのオペレーティングシステム及びネットワークソフトウェアの動作とは無関係に、並列計算機システムを構成する複数のノードの運用管理を管理端末装置で一括して行うことが可能である。

【0034】(2) 前記(1)の並列計算機システムの管理装置において、前記複数のノードのサブプロセッサは、当該ノードの主電源を投入または切断する機能を備え、前記管理端末装置は、前記複数のノードのサブプロセッサに、一括または個別に主電源を投入または切断するシステム制御コマンドを送信する手段を備えるものである。

【0035】前記(2)の並列計算機システムの管理装置では、前記メインプロセッサは主電源により動作するので、前記サブプロセッサにより前記主電源の投入または切断を行うことにより、前記メインプロセッサへの電源の投入を制御することができる。

10

20

30

40

50

【0036】また、前記管理端末装置は、前記主電源を投入または切断するシステム制御コマンドを、送信先を全てのノードまたは特定のノードに指定したバケットとして、前記システム制御インタフェースを介して、前記複数のノードに一括または個別に送信する。

【0037】前記(2)の並列計算機システムの管理装置では、前記複数のノードのサブプロセッサ及びシステム制御機構は、補助電源により動作しているので、メインプロセッサに主電源が投入されていない場合でも、動作可能である。

【0038】以上の様に、前記並列計算機システムの管理装置によれば、管理端末装置からの指示により複数のノードの主電源の投入または切断を行うので、並列計算機システムを構成する複数のノードの電源の投入または切断を管理端末装置で一括または個別に行うことが可能である。

【0039】(3)前記(2)の並列計算機システムの管理装置において、前記管理端末装置は、前記複数のノードの主電源を個別に投入するシステム制御コマンドを、予め設定された時間間隔で、前記複数のノードのサブプロセッサに個別に送信する手段を備えるものである。

【0040】前記並列計算機システムを構成する複数のノードの主電源を一齐に投入すると、前記主電源に電力を供給する電源設備に過大な突入電流が流れ、前記電源設備に負担をかけるので、前記(3)の並列計算機システムの管理装置では、前記複数のノードの主電源の投入時刻を、各ノードごとにずらし、前記電源設備の突入電流を低く抑える様にする。

【0041】これは、前記管理端末装置から予め設定された時間間隔で、前記主電源を投入するシステム制御コマンドを、送信先を特定のノードに指定したバケットとして前記システム制御インタフェースを介して送ることにより行われる。

【0042】以上の様に、前記並列計算機システムの管理装置によれば、前記複数のノードへの主電源の投入指示を、予め設定された特定の時間間隔で行うので、並列計算機システムに電力を供給する電源設備の突入電流を低く抑えることが可能である。

【0043】(4)前記(1)の並列計算機システムの管理装置において、前記管理端末装置は、前記複数のノードの特定のノードのサブプロセッサに特定のシステム制御コマンドを送信し、予め設定された時間内に前記特定のシステム制御コマンドに対する正常な応答が受信されない場合に、前記特定のノードに異常が発生しているとみなす手段を備えるものである。

【0044】前記(4)の並列計算機システムの管理装置では、前記特定のシステム制御コマンドを、前記管理端末装置から前記システム制御インタフェースを介して前記サブプロセッサに送信し、前記の送信された特定の

システム制御コマンドを前記サブプロセッサで実行した場合に、前記メインプロセッサの異常により前記特定のシステム制御コマンドの実行結果が得られない場合がある。

【0045】前記の様な場合に、前記(4)の並列計算機システムの管理装置の管理端末装置は、予め設定された時間の間、前記特定のシステム制御コマンドに対する応答を待ち、前記の予め設定された時間内に前記特定のシステム制御コマンドが正常に実行されたことを示す応答が受信されない場合に、前記特定のノードに異常が発生しているとみなす。

【0046】以上の様に、前記並列計算機システムの管理装置によれば、管理端末装置からの特定のシステム制御コマンドに対する正常なレスポンスが一定時間中に受信されるかどうかを調べるので、並列計算機システムを構成する複数のノードが正常に動作中であるかを管理端末装置で監視することが可能である。

【0047】(5)前記(1)の並列計算機システムの管理装置において、前記複数のノードのシステム制御機構は、当該ノードのメインプロセッサまたはサブプロセッサが動作時に出力するメッセージであるノードメッセージを蓄積する手段を備え、前記管理端末装置は、当該ノードのシステム制御機構に蓄積されたノードメッセージを読み取る手段を備えるものである。

【0048】前記並列計算機システムを構成する複数のノードのメインプロセッサは、各処理の段階で種々のノードメッセージを出力する。

【0049】例えば、前記並列計算機システムを構成する複数のノードのメインプロセッサは、システム立ち上げ処理中にファイルシステム上に矛盾を発見すると、特定のノードメッセージを出力し、そのファイルシステムの修復を開始する。前記メインプロセッサがファイルシステムの修復に失敗すると、前記ファイルシステムの修復に失敗したことを示すノードメッセージを出力し、前記システム立ち上げ処理は異常終了する。

【0050】また、前記並列計算機システムを構成する複数のノードのメインプロセッサは、システム立ち上げ処理が正常終了した後、動作中に回復不能な障害を検出すると、パニックメッセージと呼ばれる障害内容や障害発生箇所等の内容を含んだノードメッセージを出力し、前記回復不能な障害を検出したメインプロセッサは、通常、前記パニックメッセージを特定のディスプレイ装置に出力した直後に、システムダウンを起こして動作を停止する。

【0051】前記の様な場合には、前記ノードメッセージの内容を検討し、システム立ち上げ処理の異常終了やシステムダウンの原因を取り除く必要があるが、当該メインプロセッサは既に動作を停止しているので、従来の並列計算機システムの管理装置では、前記ノードメッセージを管理端末装置に送信して表示することはできな

10

20

30

40

50



った。

【0052】そこで、前記(5)の並列計算機システムの管理装置では、当該ノードのメインプロセッサまたはサブプロセッサが動作時に出力するノードメッセージを前記複数のノードのシステム制御機構に蓄積し、前記管理端末装置が、当該ノードのシステム制御機構に蓄積されたノードメッセージを読み取ることによって、前記ノードメッセージを前記管理端末装置に表示し、前記ノードメッセージの内容を前記管理端末装置にて検討することを可能にしている。

【0053】以上の様に、前記並列計算機システムの管理装置によれば、特定のノードのメインプロセッサまたはサブプロセッサが動作時に出力するノードメッセージを蓄積し、前記管理端末装置が前記の蓄積されたノードメッセージを読み取るので、並列計算機システムを構成する複数のノードのメインプロセッサがノードメッセージを出力した後にその動作を停止した場合であっても、前記ノードメッセージを管理端末装置で一括して管理することが可能である。

【0054】(6)前記(1)の並列計算機システムの管理装置において、前記複数のノードのサブプロセッサは、当該ノードのメインメモリまたはレジスタの内容を参照及び更新する手段を備え、前記管理端末装置は、前記複数のノードのサブプロセッサに、当該ノードのメインメモリまたはレジスタの内容を参照または更新するシステム制御コマンドを送信する手段を備えるものである。

【0055】前記並列計算機システムを構成する複数のノードのメインプロセッサに障害が発生したときに、当該メインプロセッサに接続されたメインメモリやレジスタの内容を参照して障害の原因を調べ、また、可能な場合には、前記メインメモリやレジスタの内容を変更して前記障害により中断している処理を続行したい場合がある。

【0056】前記の様な場合、発生した障害によってはメインプロセッサは正常に動作できないことがあるので、前記管理端末装置からのシステム制御コマンドにより、前記障害が発生したメインプロセッサを備えるノードのサブプロセッサが、前記メインメモリまたはレジスタの内容を参照または更新する。

【0057】以上の様に、前記並列計算機システムの管理装置によれば、管理端末装置からの指示によりノードのメインメモリまたはレジスタの内容を参照または更新するので、並列計算機システムを構成する複数のノードの障害発生時のメインメモリ及びレジスタの内容を管理端末装置で一括して管理することが可能である。

【0058】(7)前記(1)の並列計算機システムの管理装置において、前記複数のノードのサブプロセッサは、当該ノードのメインプロセッサをリセットする手段を備え、前記管理端末装置は、当該ノードのサブプロセ

ッサに、当該ノードのメインプロセッサをリセットするシステム制御コマンドを送信する手段を備えるものである。

【0059】前記並列計算機システムを構成する複数のノードにおいて、オペレーティングシステムや他のソフトウェアをバージョンアップしたり、また、障害の原因を取り除く作業を行った後等、メインプロセッサをリセットする必要がある場合がある。

【0060】前記の様な場合に、前記(7)の並列計算機システムの管理装置では、前記管理端末装置からのシステム制御コマンドにより、前記複数のノードのサブプロセッサは、当該ノードのメインプロセッサをリセットする。

【0061】以上の様に、前記並列計算機システムの管理装置によれば、管理端末装置からの指示により前記複数のノードのメインプロセッサのリセットを行うので、並列計算機システムを構成する複数のノードのメインプロセッサのリセットを管理端末装置から一括して行うことが可能である。

【0062】(8)前記(1)の並列計算機システムの管理装置において、前記複数のノードのサブプロセッサは、当該ノードのメインメモリの内容を参照及び更新する手段と、当該ノードのメインプロセッサをリセットする手段とを備え、前記管理端末装置は、当該ノードのメインプロセッサが格納しているメインメモリ中のブートストラップデバイス名を参照及び更新するシステム制御コマンドと、当該ノードのメインプロセッサをリセットするシステム制御コマンドとを送信する手段を備えるものである。

【0063】前記並列計算機システムを構成する複数のノードにおいて、あるメインプロセッサのシステム立ち上げ処理を行うときに、前記システム立ち上げ処理中にメインメモリにロードするオペレーティングシステムや他のソフトウェアを変更する場合や、或いは、オペレーティングシステムや他のソフトウェアを格納しているブートストラップデバイスに障害が発生した場合等、前記システム立ち上げ処理で使用するブートストラップデバイスの変更が必要になることがある。

【0064】この様な場合に、前記(8)の並列計算機システムの管理装置では、前記管理端末装置により、当該ノードのメインプロセッサが格納しているメインメモリ中のブートストラップデバイス名を参照するシステム制御コマンドを前記サブプロセッサに送り、前記メインメモリ中のブートストラップデバイス名を確認した後、前記管理端末装置は、当該ノードのメインプロセッサが格納しているメインメモリ中のブートストラップデバイス名を他のブートストラップデバイス名に更新するシステム制御コマンドを前記サブプロセッサに送る。

【0065】前記管理端末装置からメインメモリ中のブートストラップデバイス名を他のブートストラップデバ

イス名に更新するシステム制御コマンドを受け取った前記サブプロセッサは、当該ノードのメインメモリ中のブートストラップデバイス名を更新する。

【0066】次に、前記管理端末装置は、当該ノードのメインプロセッサをリセットするシステム制御コマンドを前記サブプロセッサに送り、当該ノードのメインプロセッサをリセットして、更新した他のブートストラップデバイスにより前記メインプロセッサのシステム立ち上げ処理を行う。

【0067】以上の様に、前記並列計算機システムの管理装置によれば、管理端末装置からの指示により前記複数のノードのメインメモリ中のブートストラップバス情報を変更し、メインプロセッサのリセットを行うので、並列計算機システムを構成する複数のノードの特定のブートストラップデバイスに障害が発生した場合に、管理端末装置からの指示により、ブートストラップデバイスを変更して前記複数のノードのメインプロセッサのシステム立ち上げ処理を行うことが可能である。

【0068】(9) 前記(1)の並列計算機システムの管理装置において、前記管理端末装置を複数備え、前記複数の管理端末装置のうちの一部の管理端末装置の機能を制限する手段を備えるものである。

【0069】前記(9)の並列計算機システムの管理装置では、複数の管理端末装置を備えることにより、特定の管理端末装置が故障した場合に、他の管理端末装置により、前記並列計算機システムの運用管理を行う。

【0070】前記の様に、前記並列計算機システムに複数の管理端末装置を接続した場合には、前記複数の管理端末装置の動作の内容及、互いに他の管理端末装置の動作の内容及と競合する場合がある。

【0071】この為、前記(9)の並列計算機システムの管理装置では、前記複数の管理端末装置が動作する場合に、特定の管理端末装置をメイン管理端末装置に、他の管理端末装置をサブ管理端末装置に設定し、サブ管理端末装置が行う動作の内容及を制限することにより、前記競合の発生を防止する。

【0072】以上の様に、前記並列計算機システムの管理装置によれば、複数の管理端末装置を備えているので、1つの管理端末装置に障害が発生した場合でも並列計算機システムの運用管理を続行し、並列計算機システムの信頼性を向上させることが可能である。

【0073】また、前記並列計算機システムの管理装置によれば、複数の管理端末装置にメイン管理端末装置とサブ管理端末装置とを設定するので、並列計算機システムを複数の管理端末装置で管理した場合に、前記複数の管理端末装置の動作の競合を防止することが可能である。

【0074】(10) 前記(1)の並列計算機システムの管理装置において、前記管理端末装置は、補助電源で動作し、特定の信号を入力すると前記管理端末装置の主

電源を投入する電源投入論理と、前記電源投入論理により主電源が投入されたときに、前記複数のノードのサブプロセッサに、一括または個別に主電源を投入するシステム制御コマンドを送信する手段を備えるものである。

【0075】前記(10)の並列計算機システムの管理装置では、前記管理端末装置に、補助電源で動作し、特定の信号を入力すると前記管理端末装置の主電源を投入する電源投入論理を接続し、前記電源投入論理をネットワークや他の通信回線に接続しておく。

【0076】また、前記管理端末装置の主電源が投入されたときに実行されるシステム立ち上げ処理の最後に、前記複数のノードのサブプロセッサに一括または個別に主電源を投入するシステム制御コマンドを送信するプログラムを追加しておく。

【0077】次に、前記ネットワークや他の通信回線を介して、他の端末装置から前記電源投入論理に特定の信号を送り、前記管理端末装置の主電源を投入する。

【0078】前記管理端末装置の主電源が投入されると、前記管理端末装置のシステム立ち上げ処理を行った後、前記複数のノードのサブプロセッサに一括または個別に主電源を投入するシステム制御コマンドを送信するプログラムが実行され、前記並列計算機システムの運用開始を、オペレータが直接前記管理端末装置を操作すること無く行うことができる。

【0079】以上の様に、前記並列計算機システムの管理装置によれば、遠隔地からのアクセスにより管理端末装置の主電源を投入するので、並列計算機システムの運用管理を遠隔地から行うことが可能である。

【0080】

【発明の実施の形態】以下、本発明について、実施形態とともに図面を参照して詳細に説明する。

【0081】なお、実施形態を説明するための全図において、同一機能を有するものは同一符号を付け、その繰り返しの説明は省略する。

【0082】(実施形態1) 以下に、本発明の並列計算機システムの管理装置を実施する実施形態1の概略構成について説明する。

【0083】図1は、本発明の並列計算機システムの管理装置を実施する実施形態1の概略構成を示す図である。図1において、100a~100cはノード、101は通信処理装置、102a~102cは通信インタフェース機構、103a~103cは通信ケーブル、104は管理端末装置、105a~105dはシステム制御機構、106a~106dは通信ケーブル、107はネットワーク集線装置、108a~108dはLAN(Local Area Network)制御機構、109a~109dは通信ケーブル、110はネットワーク集線装置である。

【0084】図1に示す様に、本実施形態の並列計算機システムの管理装置は、並列計算機システムを構成する

ノード100a~100cと、並列処理中のノード100a~100cでの通信を制御する通信処理装置101と、ノード100a~100cのシステム管理を行う管理端末装置104と、管理端末装置104とノード100a~100cとを接続するネットワーク集線装置107と、ネットワーク集線装置110とを備えており、管理端末装置104は、システム制御機構105aと、LAN制御機構108aとを有し、ノード100aは、通信インタフェース機構102aと、システム制御機構105bと、LAN制御機構108bとを有し、ノード100bは、通信インタフェース機構102bと、システム制御機構105cと、LAN制御機構108cとを有し、ノード100cは、通信インタフェース機構102cと、システム制御機構105dと、LAN制御機構108dとを有している。

【0085】また、図1に示す様に、本実施形態の並列計算機システムの管理装置では、ノード100a~100cの通信インタフェース機構102a~102cを通信ケーブル103a~103c及び通信処理装置101を介して接続し、ノード100a~100cのシステム制御機構105b~105dを通信ケーブル106a~106d及びネットワーク集線装置107を介して管理端末装置104のシステム制御機構105aに接続し、ノード100a~100cのLAN制御機構108b~108dを通信ケーブル109a~109d及びネットワーク集線装置110を介して管理端末装置104のLAN制御機構108aに接続している。

【0086】本実施形態の並列計算機システムの管理装置のシステム制御インタフェースは、前記の様に、管理端末装置104側のシステム制御機構105aとノード100a~100c側のシステム制御機構105b~105dとをイーサネット等の通信ケーブル106及びマルチポートリピータ等のネットワーク集線装置107を用いて相互接続することにより実現されるインタフェースである。

【0087】また、本実施形態の並列計算機システムの管理装置のシステム運用支援インタフェースは、管理端末装置104側のLAN制御機構108aとノード100a~100c側のLAN制御機構108b~108dとをイーサネット等の通信ケーブル109及びマルチポートリピータ等のネットワーク集線装置110を用いて相互接続することにより実現されるインタフェースである。

【0088】前記システム運用支援インタフェースは、従来の並列計算機システムの運用管理を行うインタフェースであり、ノード100a~100cのメインプロセッサが動作している場合に使用し、ノード100a~100cのメインプロセッサで実行しているアプリケーションソフトウェアが出力するメッセージを管理端末装置104に表示する等のシステム管理を行うものである。

【0089】以下に、本実施形態の並列計算機システムの管理装置において並列計算機システムを構成するノード100a~100cについて説明する。

【0090】図2は、本実施形態の並列計算機システムの管理装置において並列計算機システムを構成するノード100a~100cの概略構成を示す図である。

【0091】図2において、200は主電源、201は補助電源、202はメインプロセッサ、203はソフトウェア、204はメインメモリ、205はプロセッサメモリ制御機構、206はシステムバス、207はシステムディスク、208はI/O制御機構、209はRS-232C制御機構、210はブートストラップROM (Read Only Memory)、211はシステムサポート機構、212はサブプロセッサ、213はROM、214はSRAM (Static Random Access Memory; 不揮発メモリ)、215はローカルバス、216は電源投入/切断信号、217はプロセッサリセット信号、218はLAN制御部、219はRS-232C制御部、220はプロセッサ、221はROM、222はRAM (Random Access Memory)、223はデータインタフェース、224は制御インタフェースである。

【0092】図2に示す様に、本実施形態の並列計算機システムの管理装置のノード100a~100cは、通信インタフェース機構102a~102cと、システム制御機構105b~105dと、LAN制御機構108b~108dとを有し、ノード100a~100cで並列処理を行うアプリケーションソフトウェアを実行するメインプロセッサ202と、サブプロセッサ212を有するシステムサポート機構211と、主電源200と、補助電源201とを備えている。

【0093】また、本実施形態の並列計算機システムの管理装置のノード100a~100cは、メインプロセッサ202により実行されるオペレーティングシステム及びネットワークソフトウェアであるソフトウェア203と、ソフトウェア203を格納するメインメモリ204と、メインプロセッサ202とメインメモリ204とのインタフェース制御を行うプロセッサメモリ制御機構205と、システムバス206と、システムディスク207と、システムディスク207を制御するI/O制御機構208と、ノードメッセージの出力やシステム制御機構105b~105d経由のオペレータとのインタラクティブなやりとりを行うRS-232C制御機構209と、メインプロセッサ202のシステム立ち上げ処理を行うブートストラッププログラムを格納しているブートストラップROM210とを備えている。

【0094】本実施形態の並列計算機システムの管理装置において、サブプロセッサ212を有し、メインプロセッサ202のステータス管理等のシステム制御を行うシステムサポート機構211は、サブプロセッサ212



上で動作する制御プログラムを格納しているROM213と、ハードウェアに依存した情報を格納しているSRAM214を備えている。

【0095】本実施形態の並列計算機システムの管理装置のノード100a~100cのシステム制御機構105b~105dは、管理端末装置104との間でイーサネットパケットの送受信を制御するLAN制御部218と、RS-232C制御機構209及びサブプロセッサ212との間でのRS-232Cパケットの送受信を制御するRS-232C制御部219と、イーサネットパケットとRS-232Cパケットとのプロトコル変換を行うプロセッサ220と、プロセッサ220上で動作する制御プログラムを格納するROM221と、サブプロセッサ212及びRS-232C制御機構209から送られて来るノードメッセージを格納するRAM222とを備えている。

【0096】図2に示す様に、本実施形態の並列計算機システムの管理装置のノード100a~100cでは、システム制御機構105b~105dを、RS-232C制御部219と、データインタフェース223と、RS-232C制御機構209と、システムバス206と、プロセッサメモリ制御機構205とを介してメインプロセッサ202に接続し、また、システム制御機構105b~105dをRS-232C制御部219及び制御インタフェース224を介してシステムサポート機構211のサブプロセッサ212に接続し、システムサポート機構211のサブプロセッサ212を、ローカルバス215とプロセッサメモリ制御機構205とを介してメインプロセッサ202に接続している。また、サブプロセッサ212は、プロセッサリセット信号217によりメインプロセッサ202をリセットし、電源投入/切断信号216により主電源200を制御する。

【0097】尚、図2に示す様に、本実施形態の並列計算機システムの管理装置のノード100a~100cにおいて、システム制御機構105b~105dを、RS-232C制御部219と、データインタフェース223と、RS-232C制御機構209とを介してメインプロセッサ202に接続しているのは、システム制御機構105b~105dとメインプロセッサ202との間をRS-232C等のシリアルインタフェースで接続することによりその通信ソフトウェアをコンパクトなものとし、メインプロセッサ202に障害が発生した場合であっても、システム制御機構105b~105dとメインプロセッサ202との間の通信が、できるだけ損なわれることの無い様にする為である。

【0098】本実施形態の並列計算機システムの管理装置のノード100a~100cは、主電源200で動作する部位と補助電源201で動作する部位より構成されている。

【0099】主電源200で動作する部位としては、ノ

ード100a~100cのメインプロセッサ202、ソフトウェア203を格納するメインメモリ204、メインプロセッサ202とメインメモリ204とのインタフェース制御を行うプロセッサメモリ制御機構205、ノード100a~100cのメインプロセッサ202のシステム立ち上げ処理を行うブートストラッププログラムを格納しているブートストラップROM210等があり、これらに、システムバス206を介して、通信インタフェース機構102a~102c、LAN制御機構108b~108d等が接続され、また、システムディスク207はI/O制御機構208経由にて接続される。

【0100】補助電源201で動作する部位としては、ノード100a~100cの主電源200の制御やメインプロセッサ202のステータス管理等のシステム制御を行う部位であるシステムサポート機構211と、ノード100a~100cと管理端末装置104との通信を制御するシステム制御機構105b~105dがある。

【0101】サブプロセッサ212は、管理端末装置104からの指示により電源投入/切断信号216を出力することで、主電源200の制御を行い、プロセッサリセット信号217を出力することで、メインプロセッサ202をリセットする機能を持つ。

【0102】ノード100a~100cのノードメッセージは、メインプロセッサ202が動作し、メインプロセッサ202を制御するオペレーティングシステム及びネットワークソフトウェアであるソフトウェア203が起動されている状態では、データインタフェース223を介してRS-232C制御機構209からRAM222に蓄積され、ソフトウェア203が起動されていない状態では、サブプロセッサ212より、制御インタフェース224を介してブートストラップメッセージ等がRAM222に蓄積される。

【0103】本実施形態の並列計算機システムの管理装置のシステム制御機構105b~105dのプロセッサ220は、前記のパケットのプロトコル変換の他に、以下の処理も行う。

【0104】すなわち、管理端末装置104からのイーサネットパケットを解釈し、パケットの内容に応じた処理を行い、管理端末装置104からの指示によりRAM222に格納しているノードメッセージを管理端末装置104に送信する処理を行い、サブプロセッサ212は、制御インタフェース224を介して送られてきたパケットを解釈し、その内容に応じた制御を行う。

【0105】以下に、本実施形態の並列計算機システムの管理装置の管理端末装置104の概略構成について説明する。

【0106】図3は、本実施形態の並列計算機システムの管理装置の管理端末装置104の概略構成を示す図である。図3において、300はプロセッサ、301はソフトウェア、302はメインメモリ、303はブートス

トラップROM、304はプロセッサメモリ制御機構、305はシステムバス、306はI/O制御機構、307はシステムディスク、308、309はRS-232C制御機構、310はグラフィックス制御機構、311はLAN制御部、312はRS-232C制御部、313はプロセッサ、314はROM、315はRAM、316は制御インタフェース、317はデータインタフェースである。

【0107】図3に示す様に、本実施形態の並列計算機システムの管理装置の管理端末装置104は、管理端末装置104内の全ての処理を制御/統括するプロセッサ300と、管理端末装置104のオペレーティングシステム及びネットワークソフトウェアであるソフトウェア301が格納されているメインメモリ302と、管理端末装置104のシステム立ち上げ処理を行うブートストラッププログラムを格納しているブートストラップROM303と、プロセッサ300、メインメモリ302及びブートストラップROM303のインタフェース制御を行うプロセッサメモリ制御機構304とを備えている。

【0108】また、本実施形態の並列計算機システムの管理装置の管理端末装置104は、システムバス305と、システムディスク307を制御するI/O制御機構306と、システムディスク307と、ソフトウェア301がノード100a~100cに対し電源制御等のシステム制御コマンドを発行する際に使用するRS-232C制御機構308と、ノードメッセージの出力やシステム制御機構105a経由にてオペレータとのインタラクティブなやりとりを行うRS-232C制御機構309と、ディスプレイターミナルやキーボード及びマウスといったマンマシンインタフェースを制御するグラフィックス制御機構310と、システム制御機構105aとを備えている。

【0109】本実施形態の並列計算機システムの管理装置の管理端末装置104のシステム制御機構105aは、ノード100a~100cとの間でイーサネットパケットの送受信を制御するLAN制御部311と、RS-232C制御機構308及び309との間でのRS-232Cパケットの送受信を制御するRS-232C制御部312と、イーサネットパケットとRS-232Cパケットとのプロトコル変換を行うプロセッサ313と、プロセッサ313で動作する制御プログラムを格納するROM314と、ノード100a~100cより送られてくるノードメッセージを格納するRAM315とを備えている。

【0110】また、図3に示す様に、本実施形態の並列計算機システムの管理装置の管理端末装置104では、プロセッサ300をプロセッサメモリ制御機構304を介してメインメモリ302、ブートストラップROM303及びシステムバス305に接続し、システムディス

ク307をI/O制御機構306を介してシステムバス305に接続し、LAN制御機構108aと、RS-232C制御機構308及び309と、グラフィックス制御機構310とをシステムバス305に接続している。

【0111】更に、図3に示す様に、本実施形態の並列計算機システムの管理装置の管理端末装置104では、システム制御機構105aのRS-232C制御部312を、制御インタフェース316及びデータインタフェース317を介してRS-232C制御機構308及び309に接続している。

【0112】本実施形態の並列計算機システムの管理装置において、システム制御インタフェースは、ノード100a~100cのシステム制御機構105b~105dと管理端末装置104のシステム制御機構105aとをイーサネットケーブル等を用いて、相互接続することにより形成されている。

【0113】前記システム制御インタフェースは、管理端末装置104側のシステム制御機構105aが動作可能な状態であり、ノード100a~100cの補助電源201が投入されており、サブプロセッサ212及びシステム制御機構105b~105dが動作可能な状態であれば、ノード100a~100cの主電源200が投入されておらず、すなわちメインプロセッサ202が動作しておらず、メインプロセッサ202全体を制御するオペレーティングシステム及びネットワークソフトウェアであるソフトウェア203が起動されていなくとも使用可能である。

【0114】これに対し、システム運用支援インタフェースは、管理端末装置104のLAN制御機構108aとノード100a~100cのLAN制御機構108b~108dとをイーサネットケーブル等を用いて、相互接続することにより形成されており、前記システム運用支援インタフェースは、TCP/IP (Transmission Control Protocol/Internet Protocol) にて使用するため、管理端末装置104及びノード100a~100cのオペレーティングシステム及びそのネットワークソフトウェアであるソフトウェア203及びソフトウェア301が起動され、TCP/IPをサポートするネットワークソフトウェアを実行している状態でのみ使用可能となる。

【0115】以下に、本実施形態の並列計算機システムの管理装置の管理端末装置104とノード100a~100cとの通信シーケンスについて説明する。

【0116】図4は、本実施形態の並列計算機システムの管理装置の管理端末装置104とノード100a~100cとの通信シーケンスの一例を示す図である。図4において、401はアダプタ制御コマンド及びそのレスポンス、402はシステム制御コマンド及びそのレスポンス、403はノードメッセージである。

【0117】図4に示す様に、本実施形態の並列計算機システムの管理装置では、アダプタ制御コマンド及びそのレスポンス401、または、システム制御コマンド及びそのレスポンス402であるパケットの送受信、或いは、ノードメッセージ403の送受信により通信を行う。

【0118】アダプタ制御コマンド及びそのレスポンス401は、管理端末装置104のソフトウェア301が管理端末装置104のシステム制御機構105aと通信を行う際、およびサブプロセッサ212がシステム制御機構105b~105dと通信を行う際に使用し、制御インタフェース316または制御インタフェース224を介して送受信される。

【0119】システム制御コマンド及びそのレスポンス402は、管理端末装置104のソフトウェア301がノード100a~100cのサブプロセッサ212と通信を行う際に使用し、制御インタフェース316及び制御インタフェース224を介して送受信される。

【0120】ノードメッセージ403は、ソフトウェア203が起動していないときは、サブプロセッサ212からシステム制御機構105b~105dのRAM222へ送信されて蓄積され、また、ソフトウェア203が起動されているときは、メインプロセッサ202からRS-232C制御機構209よりシステム制御機構105b~105dのRAM222へ送信されて蓄積される。

【0121】システム制御機構105b~105dのRAM222に蓄積されたノードメッセージ403は、管理端末装置104からの要求により、ノード100a~100cのシステム制御機構105b~105dのRAM222から、管理端末装置104のシステム制御機構105aを介し、管理端末装置104のRS-232C制御機構309へ送信され、管理端末装置104のグラフィックス制御機構310に接続されるグラフィックスディスプレイ等に表示される。

【0122】以下に、本実施形態の並列計算機システムの管理装置におけるアダプタ制御コマンド及びそのレスポンス401のパケットフォーマットについて説明する。

【0123】図5は、本実施形態の並列計算機システムの管理装置におけるアダプタ制御コマンド及びそのレスポンス401のパケットフォーマットを示す図である。図5において、501は種別フィールド、502は送信元アドレスフィールド、503は受信先アドレスフィールド、504は情報部フィールド、505は識別子である。

【0124】図5に示す様に、本実施形態の並列計算機システムの管理装置におけるアダプタ制御コマンド及びそのレスポンス401のパケットは、種別フィールド501と、送信元アドレスフィールド502と、受信先ア

ドレスフィールド503と、情報部フィールド504と、識別子505とを備えている。

【0125】本実施形態の並列計算機システムの管理装置において、種別フィールド501にはアダプタ制御コマンドまたはそのレスポンスであることを示すパケット識別子、例えば「A」が格納され、送信元アドレスフィールド502にはパケットの送信元アドレス、受信先アドレスフィールド503にはパケットの受信先アドレスが格納される。

【0126】また、情報部フィールド504には、パケットの種類により、異なったパラメータが格納され、さらにパケットの末尾には、パケットの終わりを示す識別子505、例えば「LF」（ラインフィード）が付加される。

【0127】以下に、本実施形態の並列計算機システムの管理装置におけるシステム制御コマンド及びそのレスポンス402のパケットフォーマットについて説明する。

【0128】図6は、本実施形態の並列計算機システムの管理装置におけるシステム制御コマンド及びそのレスポンス402のパケットフォーマットを示す図である。図6において、601は種別フィールド、602は送信元アドレスフィールド、603は受信先アドレスフィールド、604は情報部フィールド、605は識別子である。

【0129】図6に示す様に、本実施形態の並列計算機システムの管理装置におけるシステム制御コマンド及びそのレスポンス402のパケットは、種別フィールド601と、送信元アドレスフィールド602と、受信先アドレスフィールド603と、情報部フィールド604と、識別子605とを備えている。

【0130】本実施形態の並列計算機システムの管理装置において、種別フィールド601には、システム制御コマンドまたはそのレスポンスであることを示すパケット識別子、例えば「d」が格納され、送信元アドレスフィールド602にはパケットの送信元アドレス、受信先アドレスフィールド603にはパケットの受信先アドレスが格納される。

【0131】また、情報部フィールド604には、パケットの種類により異なったパラメータが格納され、さらにパケットの末尾には、パケットの終わりを示す識別子605、例えば「LF」が付加される。

【0132】また、本実施形態の並列計算機システムの管理装置において、管理端末装置104からの送信パケットの受信先アドレスフィールド603に16進数の「0xffffffff」が格納されると、そのパケットはブロードキャストパケットとなり、全てのノード100a~100cで受信される。

【0133】尚、本実施形態の並列計算機システムの管理装置において、「0x」が付加された数字は16進数

を示すものとする。

【0134】以下に、本実施形態の並列計算機システムの管理装置におけるシステム制御機構105a~105dの、パケットモードと非パケットモードのモード遷移について説明する。

【0135】図7は、本実施形態の並列計算機システムの管理装置におけるシステム制御機構のモード遷移を示す図である。図7において、701はパケットモード、702は非パケットモード、703は「SET-MODE」コマンドである。

【0136】図7に示す様に、本実施形態の並列計算機システムの管理装置におけるシステム制御機構は、固定長のパケットの送受信を行うパケットモード701と、不定長のノードメッセージ403の送受信を行う非パケットモード702とを備え、パケットモード701と非パケットモード702のモード遷移は、サブプロセッサ212からのアダプタ制御コマンドである「SET-MODE」コマンド703を実行することにより行う。

【0137】前記の様に、本実施形態の並列計算機システムの管理装置のシステム制御機構105a~105dの動作モードは、パケットモード701及び非パケットモード702の2種類があり、パケットモード701は、管理端末装置104と複数のノード100a~100cが通信を行う際に設定されるモードであり、非パケットモード702は、特定のノードとコネクション型通信を行い、前記特定のノードからのノードメッセージ403を管理端末装置104に表示する際に設定されるモードである。

【0138】尚、本実施形態の並列計算機システムの管理装置において、管理端末装置104及びノード100a~100cのシステム制御機構105a~105dは、補助電源201投入時にはパケットモード701にて動作するものとする。

【0139】以下に、本実施形態の並列計算機システムの管理装置におけるシステム制御機構105a~105dの非パケットモード702でのコネクション状態の遷移について説明する。

【0140】図8は、本実施形態の並列計算機システムの管理装置におけるシステム制御機構の非パケットモード702でのコネクション状態の遷移を示す図である。図8において、800はディスコネクト状態、801はウェイトコネクト状態、802はコネクト状態、803

は「SET-CONNECT」コマンド、804は管理端末装置104上のシステム制御機構105aとノード100a~100c上のシステム制御機構105b~105cとの間で行われる呼制御である。

【0141】図8に示す様に、本実施形態の並列計算機システムの管理装置におけるシステム制御機構の非パケットモード702でのコネクション状態には、相手のシステム制御機構が接続されておらずRAM222にノードメッセージ403を蓄積していない状態であるディスコネクト状態800と、相手のシステム制御機構が接続されていないがノードメッセージ403をRAM222に蓄積中である状態のウェイトコネクト状態801と、相手のシステム制御機構が接続されているコネクト状態802とがあり、前記コネクション状態の遷移は、「SET-CONNECT」コマンド803またはシステム制御機構からの呼制御804により行う。

【0142】図8に示す様に、本実施形態の並列計算機システムの管理装置において、非パケットモード702設定時には、ディスコネクト状態800、ウェイトコネクト状態801及びコネクト状態802の3つのコネクト状態を保持し、ディスコネクト状態800では、システム制御機構同士の通信は不可となり、ウェイトコネクト状態801では、相手のシステム制御機構との通信は不可であるが、ノードメッセージ403は、RAM222内に順次蓄積される。

【0143】通信を行うシステム制御機構同士がコネクト状態802にあるとき、非パケットモード702でのコネクション型通信が可能となる。

【0144】これらの状態は、「SET-CONNECT」コマンド803を発行することにより遷移する。また、相手のシステム制御機構からの呼制御804によるコネクト要求があった場合、ウェイトコネクト状態801からコネクト状態802に遷移する。

【0145】本実施形態の並列計算機システムの管理装置にて使用するアダプタ制御コマンド及びそのレスポンス401の一覧を表1に示す。表1において、情報部は情報部フィールド504に格納される情報を示しており、情報部のバイト0の数字は、パケット種別を示す番号である。

【0146】

【表1】

表 1

項番	パケット種別	用 途	情 報 部		
			バイト0	バイト1	バイト2以降
1	SET-ADDRESSコマンド	システム制御機構 105 の初期設定 (論理アドレスの設定等)	1	無し	無し
2	SET-ADDRESSレスポンス	初期設定結果応答	1	完了コード	ステータス情報
3	SET-MODEコマンド	システム制御機構 105 の動作モード 設定	3	動作モード	無し
4	SET-MODEレスポンス	動作モード設定結果の応答	3	完了コード	無し
5	SET-CONNECTコマンド	非パケットモード702 でのコネクشن 制御	5	コネクشن指示	無し
6	SET-CONNECTレスポンス	コネクشن制御結果	5	完了コード	詳細情報
7	REPORT-CONNECT コネクشن状態	コネクشن状態変更通知 (相手のシステム 制御機構 105 からのコネクشن要 求を自システム212 に伝える)	A	コネクشن状態 変化状況	無し
8	REPORT-CONNECT レスポンス	REPORT-CONNECT 受信応答	A	無し	無し

【0147】本実施形態の並列計算機システムの管理装置  
にて使用するシステム制御コマンド及びそのレスポ  
ンス402の一覧を表2に示す。表2において、情報部は  
情報部フィールド604に格納される情報を示してお

り、情報部のバイト0の数字は、パケット種別を示す番  
号である。  
【0148】  
【表2】

表 2

項番	パケット種別	用 途	情 報 部		
			バイト0	バイト1	バイト2以降
1	P-ONコマンド	ノード100の電源投入	1	無し	無し
2	P-ONレスポンス	電源投入の結果応答	1	完了コード	無し
3	P-OFFコマンド	ノード100の電源切断	2	無し	無し
4	P-OFFレスポンス	電源切断の結果応答	2	完了コード	無し
5	PROC-RESETコマンド	メインプロセッサ202リセ ット	3	無し	無し
6	PROC-RESETレスポンス	リセットの結果	3	完了コード	無し
7	STATUS-READコマンド	ノード100のステータスコ ードの読み取り	4	無し	無し
8	STATUS-READレスポンス	ステータスコードの読み取り 結果	4	完了コード	ステータス 情報
9	MS-READコマンド	メインメモリ204の内容の 読み込み	5	先頭アドレス：読み出し 長さ	
10	MS-READレスポンス	読み出したメモリの値を応答	5	完了コード	読み出した値
11	MS-WRITEコマンド	メインメモリ204への書き 込み	6	先頭アドレス：書き込み データ	
12	MS-WRITEレスポンス	メインメモリ204へ書き込 み後の値を応答	6	完了コード	書き込み後の 値
13	REG-READコマンド	レジスタの内容の読み込み	7	レジスタアドレス：読み 出し長さ	
14	REG-READレスポンス	読み出したレジスタの値を応 答	7	完了コード	読み出した値
15	REG-WRITEコマンド	レジスタへの書き込み	8	レジスタアドレス：書き 込みデータ	
16	REG-WRITEレスポンス	レジスタへ書き込み後の値を 応答	8	完了コード	書き込み後の 値

【0149】以下に、本実施形態の並列計算機システム 50 の管理装置におけるノード100a～100cのシステ



ム制御機構105b~105dのプロセッサ220の処理手順について説明する。

【0150】図9は、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのシステム制御機構105b~105dのプロセッサ220の処理手順の一部を示すフローチャートである。

【0151】図9に示す様に、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのシステム制御機構105b~105dのプロセッサ220では、ステップ900の処理にて、「SET-CO 10 NNECT」コマンドや呼制御により管理端末装置104からコネクト要求があるかどうかを調べる。

【0152】ステップ900の処理で、「SET-CO NNECT」コマンドや呼制御により管理端末装置104からのコネクト要求がある場合には、ステップ901の処理に進み、ノード100a~100cのシステム制御機構105b~105dが非パケットモード702であるかどうかをチェックする。

【0153】ステップ901の処理で、ノード100a~100cのシステム制御機構105b~105dが非 20 パケットモード702であれば、ステップ902の処理へ進み、ステップ901の処理で、ノード100a~100cのシステム制御機構105b~105dが非パケットモード702でなければ、ステップ903の処理にて、サブプロセッサ212からのシステム制御コマンド「SET-MODE」により、ノード100a~100cのシステム制御機構105b~105dを非パケットモード702に設定し、ステップ902の処理へ進む。

【0154】ステップ902の処理では、ノード100a~100cのシステム制御機構105b~105dの 30 RAM222に蓄積されたノードメッセージ403をシステム制御インタフェース経由で管理端末装置104へ送信し、ステップ900の処理に戻る。

【0155】ステップ900の処理にて「SET-CO NNECT」コマンドや呼制御により管理端末装置104からコネクト要求が無い場合には、ステップ904の処理に進み、ステップ904の処理にて、システム制御コマンドにより、管理端末装置104からのシステム制御があるかどうかを調べる。

【0156】ステップ904の処理にて、前記システム 40 制御コマンドにより、管理端末装置104からのシステム制御がある場合には、ステップ905の処理に進み、ノード100a~100cのシステム制御機構105b~105dがパケットモード701かどうかをチェックする。

【0157】ステップ904の処理にて、システム制御コマンドによる管理端末装置104からのシステム制御がない場合には、ステップ909の処理に進む。

【0158】ステップ905の処理で、ノード100a~100cのシステム制御機構105b~105dがパ 50

ケットモード701であれば、ステップ906の処理へ進み、ノード100a~100cのシステム制御機構105b~105dがパケットモード701でなければ、ステップ907の処理にて、サブプロセッサ212からのシステム制御コマンド「SET-MODE」により、システム制御機構105b~105dをパケットモード701に設定し、ステップ906の処理へ進む。

【0159】ステップ906の処理にて、前記システム制御コマンドの受信先アドレスフィールド603をチェックし、前記システム制御コマンドの受信先アドレスフィールド603が、自論理アドレスまたは「0xfffffff」である場合は、ステップ908の処理に進み、前記システム制御コマンドの内容をサブプロセッサ212に通知し、ステップ900の処理に戻る。

【0160】ステップ906の処理にて、前記システム制御コマンドの受信先アドレスフィールド603が、自論理アドレス及び「0xfffffff」でない場合は、ステップ900の処理に戻る。

【0161】ステップ909の処理にて、ノード100a~100cのサブプロセッサ212からの処理の結果が返ってきたかどうかを調べ、サブプロセッサ212からの処理の結果が返ってきた場合には、ステップ910の処理に進み、管理端末装置104に対し、前記システム制御コマンドのレスポンスパケットを送信し、ステップ900の処理に戻る。

【0162】以下に、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのシステムサポート機構211のサブプロセッサ212の処理手順について説明する。

【0163】図10は、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのシステムサポート機構211のサブプロセッサ212の処理 30 手順の一部を示すフローチャートである。

【0164】図10に示す様に、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのシステムサポート機構211のサブプロセッサ212では、補助電源201が投入されると、ステップ1000の処理にて、ノード100a~100cの論理アドレスを設定し、ノード100a~100cに備えられたパネルに表示するステータスコードを格納するSRAM 214内のパネルステータス管理領域に「0000」を設定する。

【0165】次に、ステップ1001の処理にて、ノード100a~100cのシステム制御機構105b~105dを非パケットモード702に設定し、ステップ1002の処理にて、ノード100a~100cのシステム制御機構105b~105dの非パケットモード702のコネクション状態をウェイトコネクト状態801に設定する。

【0166】ノード100a~100cのシステム制御

機構105b~105dのモードを非パケットモード702に設定し、システム制御機構105b~105dの非パケットモード702のコネクション状態をウェイトコネクト状態801に設定するのは、ノード100a~100cのノードメッセージをシステム制御機構105b~105dのRAM222に蓄積すると共に、管理端末装置104のシステム制御機構105aからの呼び制御804によるコネクト要求があったときに、ノード100a~100cのシステム制御機構105b~105dのRAM222に蓄積したノードメッセージを管理端末装置104に送る為である。

【0167】また、こうすることでノード100a~100c上のソフトウェア203が起動されていない場合でも管理端末装置104からRAM222に蓄積したノードメッセージを読み出すことが可能となる。

【0168】次に、管理端末装置104のシステム制御機構105aから、ノード100a~100cのシステム制御機構105b~105dにシステム制御コマンドが送られた場合には、前記システム制御コマンドをノード100a~100cのサブプロセッサ212に送り、サブプロセッサ212にて前記システム制御コマンドを実行する。

【0169】ステップ1003の処理にて、管理端末装置104のシステム制御機構105aからノード100a~100cのシステム制御機構105b~105dを介して、ノード100a~100cの主電源200を投入または切断する電源制御指示のシステム制御コマンドが送られてきているかどうかを調べる。

【0170】ステップ1003の処理で管理端末装置104からの電源制御指示があるかどうかを調べた結果、管理端末装置104からの電源制御指示がある場合には、ステップ1004の処理にて、ノード100a~100cの主電源200を投入または切断する電源制御処理を実行し、ステップ1005の処理にて、前記電源制御処理の実行結果をノード100a~100cのシステム制御機構105b~105dへ報告した後、ステップ1003の処理に戻る。

【0171】ステップ1003の処理で管理端末装置104からの電源制御指示があるかどうかを調べた結果、管理端末装置104からの電源制御指示がない場合には、ステップ1006の処理に進み、管理端末装置104のシステム制御機構105aからノード100a~100cのシステム制御機構105b~105dを介して、ノード100a~100cに備えられたパネルを制御するパネル制御指示のシステム制御コマンドが送られてきているかどうかを調べる。

【0172】ステップ1006の処理にて、管理端末装置104からのパネル制御指示があるかどうかを調べた結果、管理端末装置104からのパネル制御指示がある場合には、ステップ1007の処理に進み、パネル制御

処理を実行し、ステップ1008の処理にて、前記パネル制御処理の実行結果をノード100a~100cのシステム制御機構105b~105dへ報告した後、ステップ1003の処理に戻る。

【0173】ステップ1006の処理にて、管理端末装置104からのパネル制御指示があるかどうかを調べた結果、管理端末装置104からのパネル制御指示がない場合には、ステップ1009の処理に進み、管理端末装置104のシステム制御機構105aからノード100a~100cのシステム制御機構105b~105dを介して、ノード100a~100cのメインプロセッサ202をリセットするリセット指示のシステム制御コマンドが送られてきているかどうかを調べる。

【0174】ステップ1009の処理にて、管理端末装置104からのリセット指示があるかどうかを調べた結果、管理端末装置104からのリセット指示がある場合には、ステップ1010の処理に進み、ノード100a~100cのメインプロセッサ202のリセット処理を実行し、ステップ1011の処理にて、前記リセット処理の実行結果をノード100a~100cのシステム制御機構105b~105dへ報告した後、ステップ1003の処理に戻る。

【0175】ステップ1012の処理にてシステム制御機構105b~105dからモード切り替えの要求があるかどうかを調べた結果、モード切り替え要求がある場合には、ステップ1013の処理に進み、アダプタ制御コマンド「SET-MODE」を実行し、システム制御機構105b~105dの動作モードを切り替え、ステップ1003の処理に戻る。

【0176】以上説明した様に、本実施形態の並列計算機システムの管理装置によれば、ノード100a~100cの補助電源201で動作し、メインプロセッサ202が使用するネットワークソフトウェア及び通信ケーブル109b~109dとは別のネットワークソフトウェア及び通信ケーブル106b~106dを使用して管理端末装置104と通信を行うシステム制御機構105b~105dに、管理端末装置104からシステム制御コマンドを送信し、前記システム制御コマンドを補助電源201で動作するサブプロセッサ212で実行することにより、複数のノード100a~100cのメインプロセッサ202の制御を行うので、並列処理を実行するメインプロセッサ202の動作並びにメインプロセッサ202のオペレーティングシステム及びネットワークソフトウェアであるソフトウェア203の動作とは無関係に、並列計算機システムを構成する複数のノード100a~100cの運用管理を管理端末装置104で一括して行うことが可能である。

【0177】(実施形態2) 以下に、本発明の並列計算機システムの管理装置において、管理端末装置104から複数のノード100a~100dに主電源200の投

入を指示し、ノード100a~100dのステータスコードを監視し、ノード100a~100dのメインプロセッサ202が動作を開始したかどうかを管理する実施形態2について説明する。

【0178】図11は、本実施形態の並列計算機システムの管理装置における管理端末装置104からノード100a~100dへ主電源200の投入を指示する電源投入シーケンスの一例を示す図である。図11において、100dはノード、1101~1112は電源投入の各段階を示すシーケンスである。

【0179】図11に示す様に、本実施形態の並列計算機システムの管理装置における管理端末装置104からノード100a~100dへ主電源200の投入を指示する電源投入シーケンスでは、シーケンス1101にて、ノード100a~100dの補助電源201が投入されている。

【0180】ノード100a~100dの補助電源201が投入されると、ノード100a~100dのサブプロセッサ212は、シーケンス1102にて、システムサポート機構211内の初期化を行い、アダプタ制御コマンド「SET-ADDRESS」によって、システム制御機構105b~105dの初期化、及び、管理端末装置104がノード100a~100dを管理するために必要なアドレスである論理アドレスの設定を行う。

【0181】ここで、例えば、論理アドレス「0x00000001」を設定する「SET-ADDRESS」コマンド及びそのレスポンスのフォーマットの一例は、下記の通りとなる。

【0182】＜コマンド＞:

A0x00000001: (受信先アドレスフィールド503は省略): 0x01 LF

＜レスポンス＞:

A0x00000001: 0x00000001: 0x01 (ステータス情報) LF  
シーケンス1103にて、管理端末装置104の電源が投入されると、管理端末装置104のブートストラップROM303に格納されているブートストラッププログラムが、管理端末装置104のシステム立ち上げ処理を行う。

【0183】シーケンス1104にて、管理端末装置104のシステム立ち上げ処理が終わると、シーケンス1105にて、管理端末装置104のソフトウェア301は、管理端末装置104の論理アドレスを「SET-ADDRESS」にて設定する。

【0184】管理端末装置104及びノード100a~100dの論理アドレスが設定されると、シーケンス1106にて、管理端末装置104のソフトウェア301は、システム制御コマンドのブロードキャストパケットを用いて、ノード100a~100dの状態を示すステータスコードを読み出す。

【0185】ステータスコードは、ノード100a~1

00dのSRAM214内のパネルステータス管理領域にて管理されており、例えば、ノード100a~100dの補助電源201が正常に投入されると、ある一定のステータスコードが前記パネルステータス管理領域に書き込まれ、また、そのステータスコードは、サブプロセッサ212により読み出すことができる(本実施形態の並列計算機システムの管理装置ではコード「0000」が読み出せるものとする。)

【0186】ここでは、管理端末装置104は「STATUS-READ」コマンドを使用して、ノード100a~100dに対し、ブロードキャストを行う。

【0187】論理アドレスが「0xa0000000」である管理端末装置104が、「STATUS-READ」コマンドをブロードキャストした場合と、そのコマンドに対する、論理アドレスが「0x00000005」であるノードからのレスポンスのフォーマットの一例は、下記の通りとなる。

【0188】＜コマンド＞:

d0xa0000000: 0xffffffff: 0x4 LF

＜レスポンス＞

d0x00000005: 0xa0000000: 0x04 0000 LF

シーケンス1107にて、ノード100a~100dで前記「STATUS-READ」コマンドが受信され、サブプロセッサ212によりステータスコード「0000」が読み出された後、シーケンス1108にて、ノード100a~100dから管理端末装置104に対し、前記の様にレスポンスが返ってくる。

【0189】ここで、管理端末装置104のソフトウェア301は、正常なレスポンスが返ってきたノードの論理アドレスと、予め管理端末装置104のソフトウェア301内または特定のファイルに保持しておいた、並列計算機システムを構成するノード100a~100dの構成情報とを照らし合わせ、正常なレスポンスが返ってこないノードに対しては、予め設定された一定の時間間隔で再び「STATUS-READ」コマンドを送るリトライ処理を行う。

【0190】シーケンス1109にて、管理端末装置104のソフトウェア301は、シーケンス1108で正常なレスポンスパケットが返ってきたノードの主電源200を「P-ON」コマンドにて投入する。

【0191】例えば、論理アドレスが「0xa0000000」である管理端末装置104から、論理アドレスが「0x00000005」であるノードに対する「P-ON」コマンド及びそのレスポンスのフォーマットの一例は、下記の通りとなる。

【0192】＜コマンド＞:

d0xa0000000: 0x00000005: 0x01LF

＜レスポンス＞:

d0x00000005: 0xa0000000: 0x01 (完了コード) LF

このとき、管理端末装置104のソフトウェア301の

制御により、予め設定された一定の時間間隔で「P-ON」コマンドをずらしながらノード100a~100dに送信することで、並列計算機システム全体に電源を供給している電源設備への突入電流を低く抑えることが出来る。

【0193】シーケンス1110にて、「P-ON」コマンドを受け取ったノード100a~100dのサブプロセッサ212は、電源投入信号216を出力し、主電源200をオンにした後、「P-ON」コマンドに対するレスポンスを、管理端末装置104に返送する。

【0194】ノード100a~100dの主電源200\*

10

\*がオンになると、メインプロセッサ202によりブートストラップROM210に格納されているブートストラッププログラムが実行され、システム立ち上げ処理が開始される。

【0195】尚、システム立ち上げ処理中にブートストラッププログラムがインクリメントするステータスコードには、例えば以下のようなものがある。ここで、本実施形態の並列計算機システムの管理装置では、ステータスコードは16進数で示されている。

【0196】

【表3】

表3

ステータスコード	意味
1000	ハードウェアの初期化中
1FFF	ハードウェア障害
2000	プログラムのロード中
2FFF	プログラムロードエラー
3000	ハードウェアブートストラップ終了
A000	ソフトウェアの初期化中
F000	アプリケーションソフトウェア動作中

【0197】ノード100a~100dのメインプロセッサ202のブートストラッププログラムは、ノード100a~100dのSRAM214内のパネルステータス管理領域にステータスコードを書き込み、システム立ち上げ処理が進むと、定期的に前記ステータスコードを更新する。

【0198】また、前記パネルステータス管理領域は、ノード100a~100dのサブプロセッサ212から

30

も参照可能であり、例えば、ノード100a~100dに備えられたパネル等の表示装置に表示することにより、オペレータに対し、前記ステータスコードを開示することも可能である。

【0199】管理端末装置104のソフトウェア301

は、これらのノード100a~100dのステータスコードを「STATUS-READ」コマンドを使用して

定期的に読み出すことにより、ノード100a~100dの状態を監視する。

【0200】シーケンス1110にて、管理端末装置104のソフトウェア301は、システム制御コマンドの送信からそのレスポンスの受信までを一定の時間で監視しており、図11に示す様に、何らかの障害が発生しており、一定時間内に正常なレスポンスが返ってこないノード100dに対しては、シーケンス1111にて、予め設定された一定の時間間隔で再度システム制御コマンドを送信するリトライ処理を行う。

正常なレスポンスが返って来なかった場合、シーケンス1112にて、管理端末装置104のソフトウェア301は、ノード100dに障害が発生していることを認識する。

【0202】管理端末装置104のソフトウェア301は、前記の様に、特定のシステム制御コマンドに対する正常なレスポンスが一定時間内の間に受信されない場合に、予め設定された一定の時間間隔で前記特定のシステム制御コマンドを再度送信する制御を行うことで、ノード100a~100dのソフトウェア203が起動されていなくとも、ノード100a~100dのメインプロセッサ202のシステム立ち上げ処理が正常に終了しているかどうかの管理を行うことが可能である。

【0203】以上説明した様に、本実施形態の並列計算機システムの管理装置によれば、管理端末装置104からの指示により複数のノード100a~100dの主電源200の投入または切断を行うので、並列計算機システムを構成する複数のノード100a~100dの主電源200の投入または切断を管理端末装置104で一括または個別に行うことが可能である。

【0204】また、本実施形態の並列計算機システムの管理装置によれば、ノード100a~100dへの主電源200の投入指示を、予め設定された特定の時間間隔で行うので、並列計算機システムに電力を供給する電源設備の突入電流を低く抑えることが可能である。

【0205】また、本実施形態の並列計算機システムの管理装置によれば、管理端末装置104からの指示によりノード100a~100dのステータスコードを読み

50

出すので、複数のノード100a~100dの状態を管理端末装置104で一括して管理することが可能である。

【0206】また、本実施形態の並列計算機システムの管理装置によれば、管理端末装置104からの特定のシステム制御コマンドに対する正常なレスポンスが一定時間中に受信されるかどうかを調べるので、並列計算機システムを構成する複数のノードが正常に動作中であるかを管理端末装置104で監視することが可能である。

【0207】(実施形態3)以下に、本発明の並列計算機システムの管理装置において、管理端末装置104にノード100aからのノードメッセージ403を表示し、必要に応じて保守を行う実施形態3について説明する。

【0208】図12は、本実施形態の並列計算機システムの管理装置における管理端末装置104にノード100aからのノードメッセージ403を表示するシーケンスの一例を示す図である。図12において、1201~1217はノードメッセージ403を表示する各段階のシーケンスを示している。

【0209】図12に示す様に、本実施形態の並列計算機システムの管理装置における管理端末装置104にノード100aからのノードメッセージ403を表示するシーケンスにおいて、シーケンス1201では、ノード100aには、予め補助電源201が投入されており、システム制御機構105b(動作モードはバケットモード701)、サブプロセッサ212及びプロセッサ220は動作可能な状態にある。

【0210】補助電源201が投入されているノード100aのサブプロセッサ212は、シーケンス1202で「SET-ADDRESS」コマンドにて、ノード100aの論理アドレスを設定する。

【0211】次に、ノード100aのサブプロセッサ212は、シーケンス1203で、「SET-MODE」コマンドにてシステム制御機構105bの動作モードを非バケットモード702(ディスコネクト状態800)に設定する。

【0212】ノード100aのサブプロセッサ212は、シーケンス1204で、さらに「SET-CONNECT」コマンドにて、コネクション状態を非バケットモード702のウェイトコネクト状態801に設定する。

【0213】一方、管理端末装置104は、シーケンス1205で、管理端末装置104の電源が投入されると、管理端末装置104のシステム立ち上げ処理を開始する。

【0214】管理端末装置104のシステム立ち上げ処理が終了すると、シーケンス1206で、管理端末装置104のソフトウェア301は、ノード100aと同様にして、「SET-ADDRESS」コマンドを用いて

管理端末装置104の論理アドレスの設定を行い、シーケンス1207で、「SET-MODE」を用いて、動作モードを非バケットモード702のディスコネクト状態800に設定する。

【0215】シーケンス1208にて、管理端末装置104のソフトウェア301は、「STATUS-READ」コマンドによってノード100aのステータスコードを読み出し、ステータスコード「0000」が読み出せると、シーケンス1209にて、「P-ON」コマンドをノード100aに送信し、ノード100aの主電源200の投入を指示する。

【0216】管理端末装置104からの「P-ON」コマンドを受信し、主電源200を投入したノード100aは、ブートストラップROM210に格納されているブートストラッププログラムをメインプロセッサ202により実行し、ノード100aのシステム立ち上げ処理を行う。

【0217】このとき、ノード100aのブートストラッププログラムから出力されるノードメッセージ403は、サブプロセッサ212を経由し、ノード100aのシステム制御機構105bのRAM222に蓄積される。

【0218】管理端末装置104のソフトウェア301は、シーケンス1210で、「SET-CONNECT」コマンドにより、管理端末装置104のシステム制御機構105aのコネクション状態をコネクト状態802にすることで、ノード100aのシステム制御機構105bのRAM222に蓄積されているノード100aのメインプロセッサ202のシステム立ち上げ処理中のノードメッセージ403の監視を開始する。

【0219】「SET-CONNECT」を受けた管理端末装置104のシステム制御機構105aは、シーケンス1211で、ノード100aのシステム制御機構105bと呼制御804を行い、これを受けたノード100aのシステム制御機構105bのコネクション状態は、ウェイトコネクト状態801からコネクト状態802に遷移する。

【0220】同時にノード100aのシステム制御機構105bは、シーケンス1212で、「REPORT-CONNECT」コマンドを、ノード100aのサブプロセッサ212に発行し、管理端末装置104からのコネクト要求があったことを伝える。

【0221】このときの「REPORT-CONNECT」コマンド及びそのレスポンスのフォーマットの一例は、下記の通りとなる。尚、以下の「REPORT-CONNECT」コマンド及びそのレスポンスでは、送受信アドレスは省略されている。

【0222】<コマンド>:

A::0xA(コネクト状態変化状況)LF

<レスポンス>:



A: : 0xA LF

シーケンス1213にて、ノード100aのノードメッセージ403は、ノード100aのシステム制御機構105bが呼制御804によるコネクト要求を受け取った時点で、ノード100aのシステム制御機構105bのRAM222からLAN制御部218を経由して管理端末装置104へ送信される。

【0223】ここで、図12には特に示していないが、もし、ノード100aのメインプロセッサ202のシステム立ち上げ処理中に、ノード100aのメインプロセッサ202が使用するファイルシステムに矛盾が発見され、前記システム立ち上げ処理が中断した場合には、シーケンス1214にて、オペレータは、管理端末装置104の表示装置に出力されるノード100aのメインプロセッサ202のシステム立ち上げ処理中のノードメッセージ403により、ノード100aに障害が発生していることを認識し、UNIXのfsck等のファイルシステムを検査する保守コマンドを投入することで、ノード100aの保守を行うことも可能である。

【0224】また、ノード100aのメインプロセッサ202のシステム立ち上げ処理は正常終了したが、その後の通常の業務でノード100aのメインプロセッサ202を使用中に、パニックメッセージを出力してノード100aのメインプロセッサ202がシステムダウンを起こした場合には、オペレータは、ノード100aのシステム制御機構105bのRAM222に蓄積されたノードメッセージ403を管理端末装置104に表示させ、ノードメッセージ403の内容によりシステムダウンの要因を検討することも可能である。

【0225】ノード100aとのコネクションを切断する場合には、シーケンス1210~1212までの処理と同様、管理端末装置104のソフトウェア301が、シーケンス1215にて、「SET-CONNECT」コマンドを発行する。

【0226】管理端末装置104のソフトウェア301が「SET-CONNECT」コマンドを発行することにより、管理端末装置104のシステム制御機構105aはディスコネクト状態800になり、シーケンス1216にて、管理端末装置104のシステム制御機構105aは、ノード100aのシステム制御機構105bに対し呼制御804を行う。

【0227】シーケンス1217にて、ノード100aのシステム制御機構105bは、前記の管理端末装置104のシステム制御機構105aからの呼制御804により、相手からコネクション断の要求があったことを認識し、同時にノード100aのサブプロセッサ212に対し、このことを「REPORT-CONNECT」コマンドにて報告する。

【0228】以上の様に、管理端末装置104からノード100a~100cに対し、システム制御コマンドを

送信することで、管理端末装置104からノード100a~100cの主電源200の電源制御及びノード100a~100cの状態監視が可能となる。

【0229】以上説明した様に、本実施形態の並列計算機システムの管理装置によれば、パケットモード701及び非パケットモード702のモード切り替えをサブプロセッサ212により行うので、パケットを用い、システム制御コマンド及びそのレスポンスを複数のノード100a~100cと送受信する通信と、特定のノードとのコネクションを設定し、特定のノードのノードメッセージ403を連続して受信する通信とを、切替装置の様な特別のハードウェアを用いることなく同一の管理端末装置104で行うことが可能である。

【0230】また、本実施形態の並列計算機システムの管理装置によれば、特定のノードのメインプロセッサ202またはサブプロセッサ212が動作時に出力するノードメッセージ403を蓄積し、管理端末装置104が前記の蓄積されたノードメッセージ403を読み取るので、並列計算機システムを構成する複数のノード100a~100cのメインプロセッサ202がノードメッセージ403を出力した後にその動作を停止した場合であっても、ノードメッセージ403を管理端末装置104で一括して管理することが可能である。

【0231】（実施形態4）以下に、本発明の並列計算機システムの管理装置において、ノード100a~100cのメインプロセッサ202のシステム立ち上げ処理を実行し、必要に応じてそのブートストラップデバイスを変更してシステム立ち上げ処理を行う実施形態4について説明する。

【0232】本実施形態の並列計算機システムの管理装置では、ブートストラップデバイスからのロード処理が異常終了した場合に、ノード100a~100cのメインメモリ204の内容を読み書きすることによってその内容を変更し、ノード100a~100cのメインプロセッサ202をリセットすることによって、他のブートストラップデバイスからのロード処理を行うことが可能である。

【0233】本実施形態の並列計算機システムの管理装置において、ノード100a~100cのメインメモリ204の内容を読み書きする場合には、「MS-READ」コマンド及び「MS-WRITE」コマンドを使用する。これらのシステム制御コマンドは、ノード100a~100cのメインプロセッサ202にて通常の業務として並列処理を実行中に障害が発生したときに、ノード100a~100cのメインメモリ204の内容を調査する場合にも使用することが可能である。

【0234】また、本実施形態の並列計算機システムの管理装置では、ノード100a~100cのメインプロセッサ202、サブプロセッサ212及びプロセッサメモリ制御機構205等のハードウェアモジュール内のレ

ジスタの内容を読み書きすることも可能であり、その場合には、「REG-READ」コマンド及び「REG-WRITE」コマンドを使用する。

【0235】例えば、ノード100a~100cのメインプロセッサ202にて、通常の業務である並列処理を実行中に障害が発生した場合は、ノード100a~100c内の各ハードウェアリソースが採取する障害ログをレジスタに退避しておき、前記の「REG-READ」コマンドにより管理端末装置104から前記障害ログを読み出すことにより、管理端末装置104からの障害要因の特定が可能となる。

【0236】また、本実施形態の並列計算機システムの管理装置では、「STATUS-READ」コマンドにより、ノード100a~100cのステータスコードを読み出し、システムダウンを起こしているノードがあれば、「PROC-RESET」コマンドを送信することで、前記のシステムダウンを起こしているノードのメインプロセッサ202をリセットし、再起動させるオペレーションも可能となる。

【0237】以下に、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのメインプロセッサ202のシステム立ち上げ処理について説明する。

【0238】図13は、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのメインプロセッサ202のシステム立ち上げ処理手順を示すフローチャートである。

【0239】図14は、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのSRAM214のメモリマップを示す図である。図14において、1400はプライマリブートストラップバス情報、1401はオルタネートブートストラップバス情報である。

【0240】図14に示す様に、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのSRAM214のメモリマップは、オペレーティングシステム等のソフトウェア203を格納している第1のブートストラップデバイスを示すプライマリブートストラップバス情報1400と、第1のブートストラップデバイスが使用できない場合に使用するブートストラップデバイスを示すオルタネートブートストラップバス情報1401とを備えている。

【0241】図13に示す様に、本実施形態の並列計算機システムの管理装置におけるノード100a~100cのメインプロセッサ202のシステム立ち上げ処理手順では、管理端末装置104からの電源投入指示により主電源200が投入されると、メインプロセッサ202によりブートストラップROM210に格納されているブートストラッププログラムが実行され、ステップ1300の処理にて、SRAM214内のパネルステータス

管理領域のステータスコードを「1000」とし、ステップ1301の処理にて、ノード100a~100c内の各ハードウェアモジュールの初期診断及び初期化を行う。

【0242】ステップ1302の処理では、ステップ1301の処理のハードウェアの初期診断及び初期化が正常終了したかどうかをチェックし、ステップ1301の処理でハードウェアの初期診断及び初期化が正常終了している場合には、ステップ1303の処理に進む。

10 【0243】ステップ1301の処理でハードウェアの初期診断及び初期化が異常終了している場合には、ステップ1313の処理にて、当該ノードに備えられたパネルにステータスコード「1FFF」を表示し、当該ノードのメインプロセッサ202のシステム立ち上げ処理は異常終了となる。

【0244】ステップ1303の処理では、ステータスコードを「2000」とし、ステップ1304の処理にて、SRAM214に格納されているハードウェア依存情報のうち、図14に示すブートストラップバス情報を参照し、プライマリブートストラップバス情報1400にて指定されるブートストラップデバイス（例えば、システムディスク207等）から、オペレーティングシステム等のソフトウェア203をメインメモリ204にロードする。

【0245】SRAM214内のブートストラップバス情報は、ブートストラップROM210に格納されているブートストラッププログラムの実行時にメインメモリ204内の特定領域にコピーされ、システムが立ち上がるとソフトウェア203にて参照可能となる。

30 【0246】また、本実施形態の並列計算機システムの管理装置では、ブートストラップデバイスには、自ノード内のローカルファイルの他にイーサネット経由（システム制御インタフェース）にてbootpプロトコル（Request For Connectブートのベースとなるプロトコル）を使用し、イーサネットに接続される他のノードから取得可能となるブートストラップファイルも適用可能である。

40 【0247】ステップ1305の処理では、プライマリブートストラップバス情報1400にて指定されるブートストラップデバイスからオペレーティングシステム等のソフトウェア203をメインメモリ204にロードするロード処理が正常終了したかどうかをチェックしており、前記のロード処理に成功すると、ステップ1306の処理に進み、失敗するとステップ1314の処理に進む。

50 【0248】ステップ1306の処理にて、ステータスコードを「3000」とし、メインメモリ204にロードされたソフトウェア203が起動され、ステップ1307の処理でステータスコードを「A000」とし、ステップ1308の処理にて各種システムパラメータを設

定し、ステップ1309の処理にて、ファイルシステムの初期化を行い、ステップ1310の処理にて、TCP/IPなどのネットワークの初期化を行う。

【0249】本実施形態の並列計算機システムの管理装置では、ノード100a~100cのメインプロセッサ202で動作するオペレーティングシステム及びネットワークソフトウェア等のソフトウェア203の機能を使用するシステム運用支援インタフェースは、この時点で使用可能となる。

【0250】ステップ1311の処理にて、アプリケーションソフトウェアの起動を行い、ステップ1312の処理にてステータスコードを「F000」とし、メインプロセッサ202のシステム立ち上げ処理を終了する。

【0251】一方、ステップ1314の処理では、SRAM214内のオルタネートブートストラップバス情報1401を参照し、オルタネートブートストラップバス情報1401にて指定されるブートストラップデバイス（本実施形態の並列計算機システムの管理装置では特に開示していないが、DAT (Digital Audio Tape) 等の入出力装置）からのオペレーティングシステム等のソフトウェア203をメインメモリ204にロードする。

【0252】ステップ1315の処理にて、オルタネートブートストラップバス情報1401にて指定されるブートストラップデバイスからのロードに成功したかどうかをチェックし、成功するとステップ1306の処理に進む。

【0253】ステップ1315の処理にて、オルタネートブートストラップバス情報1401にて指定されるブートストラップデバイスからのロードが成功しない場合、ステップ1316の処理にて、オペレータによるブートストラップデバイス指定によりロード処理を行う。

【0254】ステップ1317の処理にて、ステップ1316の処理でのオペレータのブートストラップデバイス指定によるロード処理が正常終了したかどうかをチェックし、正常終了している場合にはステップ1306の処理に進み、正常終了していない場合には、ステップ1318の処理にて、ステータスコードを「2FFF」とし、メインプロセッサ202のシステム立ち上げ処理が異常終了する。

【0255】前記の様にして行ったノード100a~100cのメインプロセッサ202のシステム立ち上げ処理が異常終了した場合には、さらに、以下の様に、ブートストラップデバイスを変更したシステム立ち上げ処理を行う。

【0256】管理端末装置104のソフトウェア301は、「MS-READ」コマンドを使用して、システム制御インタフェース経由にて、ノード100a~100cのメインメモリ204のブートストラップバス情報が格納されている前記特定領域を参照し、メインプロセ

ッサ202のシステム立ち上げ処理に失敗したブートストラップデバイスを確認する。

【0257】次に、管理端末装置104のソフトウェア301は、「MS-WRITE」コマンドを使用し、システム制御インタフェース経由にて、ノード100a~100cのメインメモリ204のブートストラップバス情報が格納されている前記特定領域に、メインプロセッサ202のシステム立ち上げ処理に失敗したブートストラップデバイス以外のブートストラップデバイス名を書き込む。

【0258】管理端末装置104のソフトウェア301は、前記の様に、ノード100a~100cのメインメモリ204の前記特定領域のブートストラップバス情報を書き替えた後、「PROC-RESET」コマンドを使用し、ノード100a~100cのメインプロセッサ202をリセットしてメインプロセッサ202のシステム立ち上げ処理を再度行うことで、ブートストラップ先を変更したシステム立ち上げ処理を行うことが出来る。

【0259】また、ブートストラップバス情報の書き換えについては、ノード100a~100cのメインプロセッサ202のシステム立ち上げ処理が正常終了している場合には、以下の方法でも可能である。

【0260】すなわち、ノード100a~100cのSRAM214のブートストラップバス情報は、ノード100a~100cのソフトウェア203からも書き換え可能であるので、管理端末装置104のソフトウェア301は、システム運用支援インタフェース経由にて、ノード100a~100cのソフトウェア203に対し、ブートストラップバス情報の書き換えを指示し、指示されたソフトウェア203が当該ノードのブートストラップバス情報を書き替える。

【0261】ノード100a~100cのソフトウェア203は、更新されたブートストラップバス情報をシステム制御インタフェース経由にて管理端末装置104のソフトウェア301に通知し、管理端末装置104のソフトウェア301が、システム制御インタフェース経由にて、前記「PROC-RESET」コマンドを使用してノード100a~100cのメインプロセッサ202をリセットすれば、直ちに更新されたブートストラップバスからのロード処理が行われる。

【0262】以上説明した様に、本実施形態の並列計算機システムの管理装置によれば、管理端末装置104からの指示によりノード100a~100cのメインメモリ204またはレジスタの内容を参照または更新するので、並列計算機システムを構成する複数のノード100a~100cの障害発生時のメインメモリ204の内容を管理端末装置104で一括して管理することが可能である。

【0263】また、本実施形態の並列計算機システムの管理装置によれば、管理端末装置104からの指示によ

リノード100a~100cのメインプロセッサ202のリセットを行うので、並列計算機システムを構成する複数のノード100a~100cのメインプロセッサ202のリセットを管理端末装置104から一括して行うことが可能である。

【0264】また、本実施形態の並列計算機システムの管理装置によれば、管理端末装置104は、ノード100a~100cとの間のインタフェースを使い分けることが可能であり、管理端末装置104からの指示によりノード100a~100cのメインメモリ204のブートストラップバス情報を変更し、メインプロセッサ202のリセットを行うので、並列計算機システムを構成する複数のノード100a~100cの特定のブートストラップデバイスに障害が発生した場合に、管理端末装置104からの指示により、ブートストラップデバイスを変更してノード100a~100cのメインプロセッサ202のシステム立ち上げ処理を行うことが可能である。

【0265】(実施形態5)以下に、本発明の並列計算機システムの管理装置において、複数の管理端末装置を用いて信頼性を向上させた実施形態5の概略構成について説明する。

【0266】図15は、本発明の並列計算機システムの管理装置において、管理端末装置を二重化した実施形態5の概略構成を示す図である。図15において、105eはシステム制御機構、106eは通信ケーブル、108eはLAN制御機構、109eは通信ケーブル、111は管理端末装置である。

【0267】図15に示す様に、本実施形態の並列計算機システムの管理装置は、管理端末装置111と、通信ケーブル106eと、通信ケーブル109eとを備え、管理端末装置111は、システム制御機構105eと、LAN制御機構108eとを有しており、管理端末装置111のシステム制御機構105eを通信ケーブル106eを介してネットワーク集線装置107に接続し、管理端末装置111のLAN制御機構108eを通信ケーブル109eを介してネットワーク集線装置110に接続している。

【0268】前記の様に、本実施形態の並列計算機システムの管理装置では、複数の管理端末装置104及び111を備えているので、1つの管理端末装置が故障しても、他の管理端末装置により、並列計算機システムの運用管理を続行することが可能であるが、複数の管理端末装置を同時に使用して並列計算機システムの運用管理を行うと、複数の管理端末装置が送信するシステム制御コマンドやアダプタ制御コマンドの内容が互いに競合することがあるので、複数の管理端末装置を用いているときに管理端末装置の動作の競合を防止する処理が必要になる。

【0269】以下に、本実施形態の並列計算機システム

の管理装置において複数の管理端末装置を用いているときに管理端末装置の動作の競合を防止する処理手順について説明する。

【0270】図16は、本実施形態の並列計算機システムの管理装置において複数の管理端末装置の動作の競合を防止する処理手順を示すフローチャートである。

【0271】本実施形態の並列計算機システムの管理装置において、管理端末装置を二重化している場合には、管理端末装置の二重化情報を、例えば、管理端末装置104及び管理端末装置111の両方のソフトウェア301から参照可能な記憶領域に予め設定しておくことで、二重化した管理端末装置の競合を防止することが可能となる。

【0272】図16に示す様に、本実施形態の並列計算機システムの管理装置において管理端末装置を二重化しているときの処理手順では、ステップ1600の処理で、管理端末装置104及び管理端末装置111の両方のソフトウェア301は、管理端末装置が二重化されていることを示す二重化ビットを参照し、ビットが立っている場合には、管理端末装置が二重化されていることを認識する。

【0273】ステップ1601の処理では、ネットワーク(例えば、システム運用支援インタフェース)經由にて、相手の管理端末装置のIPアドレスを取得する。

【0274】ステップ1602の処理では、メイン管理端末装置と、前記メイン管理端末装置をバックアップするサブ管理端末装置とを決定するため、例えば、IPアドレスの若い方をメイン管理端末装置、そうでない方をサブ管理端末装置とする。

【0275】このとき、メイン管理端末装置のみを動作させておき、前記メイン管理端末装置に障害が発生したときに、直ちにサブ管理端末装置に切り替える運用方法と、メイン管理端末装置とサブ管理端末装置とを同時に動作させる運用方法とを行うことが可能であるが、後者の場合は、双方からのノード100a~100cを制御するシステム制御コマンドや、アダプタ制御コマンドの内容が競合することがあるため、サブ管理端末装置から送信可能なシステム制御コマンド及びアダプタ制御コマンドを一部制限する。

【0276】例えば、ステップ1603の処理にて、自管理端末装置がメイン管理端末装置であるかどうかを判定し、メイン管理端末装置でなかった場合には、ステップ1604の処理にて、システム制御コマンド(「P-ON」「P-OFF」等)や、また、アダプタ制御コマンド(「SET-CONNECT」等)を発行禁止にすることで、ノード100a~100cを制御するシステム制御コマンドや、アダプタ制御コマンドの内容が競合しても、並列計算機システムとしての整合性を保つことが可能である。

【0277】以上説明した様に、本実施形態の並列計算

機システムの管理装置によれば、複数の管理端末装置を備えているので、1つの管理端末装置に障害が発生した場合でも並列計算機システムの運用管理を続行し、並列計算機システムの信頼性を向上させることが可能である。

【0278】また、本実施形態の並列計算機システムの管理装置によれば、複数の管理端末装置にメイン管理端末装置とサブ管理端末装置とを設定するので、並列計算機システムを複数の管理端末装置で管理した場合に、前記複数の管理端末装置の動作の競合を防止することが可能である。

【0279】（実施形態6）以下に、本発明の並列計算機システムの管理装置において、管理端末装置104に補助電源で動作する電源投入論理を付加し、管理端末装置104の主電源を遠隔地から投入することにより並列計算機システムの主電源の投入を行う実施形態6について説明する。

【0280】図17は、本実施形態の並列計算機システムの管理装置における管理端末装置104に補助電源で動作する電源投入論理を付加した場合の管理端末装置内のハードウェアの概略構成を示す図である。図17において、1700は補助電源、1701は電源投入論理、1702は電源制御信号、1703は主電源、1704は端末装置、1705はネットワークである。

【0281】図17に示す様に、本実施形態の並列計算機システムの管理装置における管理端末装置104は、補助電源1700と、電源投入論理1701と、主電源1703とを備え、補助電源1700から電力の供給を受けている電源投入論理1701を電源制御信号1702を介して主電源1703に接続すると共にネットワーク1705を介して別の端末装置1704に接続している。

【0282】図17に示す様に、本実施形態の並列計算機システムの管理装置における管理端末装置104は、補助電源1700で動作する電源投入論理1701を設けており、電源投入論理1701は、ここでは特に図示していないが、ネットワーク制御部、電源制御部及びマイクロプロセッサ等から構成されており、ネットワーク1705経由で電源制御指示を受け取ると、主電源1703を制御する論理回路を備えている。

【0283】この電源投入論理1701により、例えば下記のような管理端末装置104の遠隔オペレーションが可能となる。

【0284】本実施形態の並列計算機システムの管理装置において、ネットワーク1705で接続された別の端末装置1704は、例えばtelnetプロトコルを使用して、電源投入論理1701にログインする。（この時、管理端末装置104には補助電源1700が投入されている状態である。）

次に、端末装置1704は、電源投入論理1701にバ

ワーオンコマンドを発行する。電源投入論理1701は、パワーオンコマンドを受け取ると、外部から電源投入指示があったことを認識し、電源制御信号1702を出力し、管理端末装置104の主電源1703を投入する。

【0285】管理端末装置104の主電源1703が投入されると、ブートストラップROM303に格納されているブートストラッププログラムが管理端末装置104のシステム立ち上げ処理を行い、ソフトウェア301を起動する。

【0286】図18は、本実施形態の並列計算機システムの管理装置における管理端末装置104のシェルプログラムの一例を示す図である。ここで、シェルプログラムとは、汎用のオペレーティングシステムであるUNIXで実行される複数のコマンド名またはプログラム名を記載した、一連の手続きを行うプログラムを指すが、図18においては、UNIXのコマンド名またはプログラム名の代わりに、そのコマンドの機能を簡単に記載している。

【0287】図18に示す様に、本実施形態の並列計算機システムの管理装置の管理端末装置104のシステム立ち上げ処理の際に実行されるシェルプログラムに、予め、ノード100a~100cの主電源200を投入指示するシステム制御コマンドを記載しておき、管理端末装置104の主電源1703が投入されたときに、このシェルプログラムが実行されるようにしておく。

【0288】このようにすることで管理端末装置104の起動を契機として、ノード100a~100cの主電源200を投入し、ノード100a~100cのメインプロセッサ202のシステム立ち上げ処理を自動的に行うことが可能である。

【0289】以上説明した様に、本実施形態の並列計算機システムの管理装置によれば、遠隔地からのアクセスにより管理端末装置104の主電源1703を投入し、さらにノード100a~100cの起動（主電源200の投入）が可能となり、並列計算機システムの運用管理を遠隔地から行うことができる。

【0290】以上、説明してきた本実施形態の並列計算機システムの管理装置では、特に図示していないが、下記のようなシステムにも適用可能である。

【0291】（1）各ノードに汎用のオペレーティングシステムを搭載していない、特定の機能を実行する専用の並列計算機システムにおいては、汎用のオペレーティングシステムのネットワーク機能を使用しない前記システム制御インタフェースのみを用いて運用管理を行う。

【0292】本発明の並列計算機システムの管理装置によれば、前記システム制御インタフェースは、運用管理の対象となるプロセッサとは独立した補助電源とネットワーク機能を備えており、汎用のオペレーティングシステムのTCP/IP等のネットワーク機能を使用しない

10

20

30

40

50



ので、前記汎用のオペレーティングシステムを搭載していない専用の並列計算機システムにおいても適用することが可能である。

【0293】(2) 各ノードに補助電源で動作する機能を持たない、或いは補助電源で動作する機能が限定されている様な並列計算機システムにおいては、補助電源を使用しない前記システム運用支援インタフェースのみを用いて管理を行う。

【0294】この場合には、主電源の投入等、補助電源を必須とする機能を除き、システム制御インタフェースの機能をシステム運用支援インタフェースによって代行することにより、本発明の並列計算機システムの管理装置を適用することが可能である。

【0295】以上、本発明を、前記実施形態に基づき具体的に説明したが、本発明は、前記実施形態に限定されるものではなく、その要旨を逸脱しない範囲において種々変更可能であることは勿論である。

【0296】

【発明の効果】本願において開示される発明のうち代表的なものによって得られる効果を簡単に説明すれば、下記のとおりである。

【0297】(1) 複数のノードの補助電源で動作し、メインプロセッサが使用するネットワークソフトウェア及び通信ケーブルとは別のネットワークソフトウェア及び通信ケーブルを使用して管理端末装置と通信を行うシステム制御機構に対し、前記管理端末装置からシステム制御コマンドを送信し、前記システム制御コマンドを前記補助電源で動作するサブプロセッサで実行することにより複数のノードのメインプロセッサの制御を行うので、並列処理を実行するメインプロセッサの動作並びに前記メインプロセッサのオペレーティングシステム及びネットワークソフトウェアの動作とは無関係に、並列計算機システムを構成する複数のノードの運用管理を管理端末装置で一括して行うことが可能である。

【0298】(2) 管理端末装置からの指示により複数のノードの主電源の投入または切断を行うので、並列計算機システムを構成する複数のノードの電源の投入または切断を管理端末装置で一括または個別に行うことが可能である。

【0299】(3) 複数のノードへの主電源の投入指示を、予め設定された特定の時間間隔で行うので、並列計算機システムに電力を供給する電源設備の突入電流を低く抑えることが可能である。

【0300】(4) 管理端末装置からの特定のシステム制御コマンドに対する正常なレスポンスが一定時間中に受信されるかどうかを調べるので、並列計算機システムを構成する複数のノードが正常に動作中であるかを管理端末装置で監視することが可能である。

【0301】(5) 特定のノードのメインプロセッサまたはサブプロセッサが動作時に出力するノードメッセ

ジを蓄積し、管理端末装置が前記の蓄積されたノードメッセージを読み取るので、並列計算機システムを構成する複数のノードのメインプロセッサがノードメッセージを出力した後にその動作を停止した場合であっても、前記ノードメッセージを管理端末装置で一括して管理することが可能である。

【0302】(6) 管理端末装置からの指示によりノードのメインメモリまたはレジスタの内容を参照または更新するので、並列計算機システムを構成する複数のノードの障害発生時のメインメモリ及びレジスタの内容を管理端末装置で一括して管理することが可能である。

【0303】(7) 管理端末装置からの指示により複数のノードのメインプロセッサのリセットを行うので、並列計算機システムを構成する複数のノードのメインプロセッサのリセットを管理端末装置から一括して行うことが可能である。

【0304】(8) 管理端末装置からの指示により複数のノードのメインメモリ中のブートストラップバス情報を変更し、メインプロセッサのリセットを行うので、並列計算機システムを構成する複数のノードの特定のブートストラップデバイスに障害が発生した場合に、管理端末装置からの指示により、ブートストラップデバイスを変更して前記複数のノードのメインプロセッサのシステム立ち上げ処理を行うことが可能である。

【0305】(9) 複数の管理端末装置を備えることも可能であるので、1つの管理端末装置に障害が発生した場合でも並列計算機システムの運用管理を続行することが可能であり、複数の管理端末装置をメイン管理端末装置とサブ管理端末装置とに設定するので、並列計算機システムを複数の管理端末装置で管理した場合に、前記複数の管理端末装置の動作の競合を防止することが可能である。

【0306】(10) 遠隔地からのアクセスにより管理端末装置の主電源を投入し、さらに複数のノードの主電源を投入するので、並列計算機システムの運用管理を遠隔地から行うことが可能である。

【図面の簡単な説明】

【図1】本発明の並列計算機システムの管理装置を実施する実施形態1の概略構成を示す図である。

【図2】実施形態1の並列計算機システムの管理装置において並列計算機システムを構成するノードの概略構成を示す図である。

【図3】実施形態1の並列計算機システムの管理装置における管理端末装置の概略構成を示す図である。

【図4】実施形態1の並列計算機システムの管理装置における管理端末装置と各ノードとの通信シーケンスの一例を示す図である。

【図5】実施形態1の並列計算機システムの管理装置におけるアダプタ制御コマンド及びそのレスポンスのパケットフォーマットを示す図である。

【図6】実施形態1の並列計算機システムの管理装置におけるシステム制御コマンド及びそのレスポンスのパケットフォーマットを示す図である。

【図7】実施形態1の並列計算機システムの管理装置におけるシステム制御機構のモード遷移を示す図である。

【図8】実施形態1の並列計算機システムの管理装置におけるシステム制御機構の非パケットモードでのコネクション状態の遷移を示す図である。

【図9】実施形態1の並列計算機システムの管理装置におけるシステム制御機構のプロセッサの処理手順の一部を示すフローチャートである。

【図10】実施形態1の並列計算機システムの管理装置におけるシステムサポート機構のサブプロセッサの処理手順の一部を示すフローチャートである。

【図11】実施形態2の並列計算機システムの管理装置における管理端末装置から各ノードへ主電源の投入を指示する電源投入シーケンスの一例を示す図である。

【図12】実施形態3の並列計算機システムの管理装置における管理端末装置に各ノードのノードメッセージを表示するシーケンスの一例を示す図である。

【図13】実施形態4の並列計算機システムの管理装置におけるノードのメインプロセッサのシステム立ち上げ処理手順を示すフローチャートである。

【図14】実施形態4の並列計算機システムの管理装置におけるノード内のSRAM内のメモリマップを示す図である。

【図15】本発明の並列計算機システムの管理装置において管理端末装置を二重化した実施形態5の概略構成を示す図である。

【図16】実施形態5の並列計算機システムの管理装置において複数の管理端末装置の動作の競合を防止する処理手順を示すフローチャートである。

【図17】実施形態6の並列計算機システムの管理装置における管理端末装置に補助電源で動作する電源投入論理を付加した場合の管理端末装置内のハードウェアの概略構成を示す図である。

【図18】実施形態6の並列計算機システムの管理装置における管理端末装置のシェルプログラムを示す。

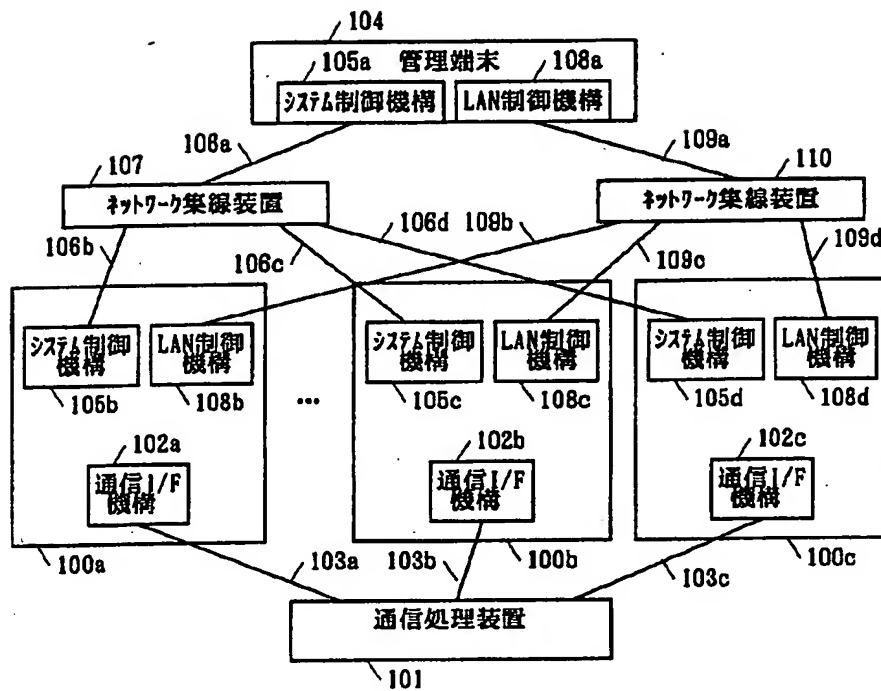
#### 【符号の説明】

100a~100c...ノード、101...通信処理装置、102a~102c...通信インタフェース機構、103a~103c...通信ケーブル、104...管理端末装置、105a~105d...システム制御機構、106a~106d...通信ケーブル、107...ネットワーク集線装

置、108a~108d...LAN制御機構、109a~109d...通信ケーブル、110...ネットワーク集線装置、200...主電源、201...補助電源、202...メインプロセッサ、203...ソフトウェア、204...メインメモリ、205...プロセッサメモリ制御機構、206...システムバス、207...システムディスク、208...I/O制御機構、209...RS-232C制御機構、210...ブートストラップROM、211...システムサポート機構、212...サブプロセッサ、213...ROM、214...SRAM、215...ローカルバス、216...電源投入/切断信号、217...プロセッサリセット信号、218...LAN制御部、219...RS-232C制御部、220...プロセッサ、221...ROM、222...RAM、223...データインタフェース、224...制御インタフェース、300...プロセッサ、301...ソフトウェア、302...メインメモリ、303...ブートストラップROM、304...プロセッサメモリ制御機構、305...システムバス、306...I/O制御機構、307...システムディスク、308及び309...RS-232C制御機構、310...グラフィックス制御機構、311...LAN制御部、312...RS-232C制御部、313...プロセッサ、314...ROM、315...RAM、316...制御インタフェース、317...データインタフェース、401...アダプタ制御コマンド及びそのレスポンス、402...システム制御コマンド及びそのレスポンス、403...ノードメッセージ、501...種別フィールド、502...送信元アドレスフィールド、503...受信先アドレスフィールド、504...情報部フィールド、505...識別子、601...種別フィールド、602...送信元アドレスフィールド、603...受信先アドレスフィールド、604...情報部フィールド、605...識別子、701...パケットモード、702...非パケットモード、703...「SET-MODE」コマンド、800...ディスコネクト状態、801...ウェイトコネクト状態、802...コネクト状態、803...「SET-CONNECT」コマンド、804...システム制御機構間の呼制御、1400...プライマリブートストラップバス情報、1401...オルタネートブートストラップバス情報、105e...システム制御機構、106e...通信ケーブル、108e...LAN制御機構、109e...通信ケーブル、111...管理端末装置、1700...補助電源、1701...電源投入論理、1702...電源制御信号、1703...主電源、1704...端末装置、1705...ネットワーク。

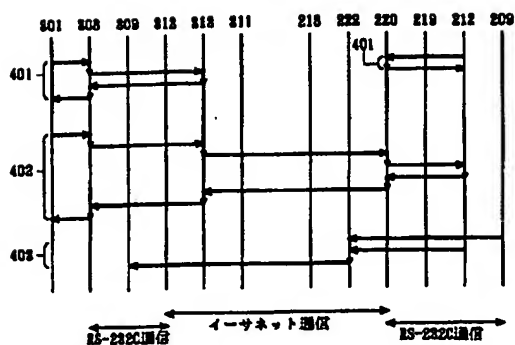
【図1】

図1



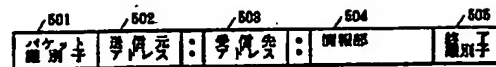
【図4】

図4



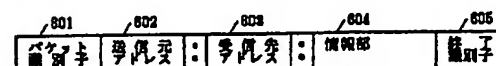
【図5】

図5



【図6】

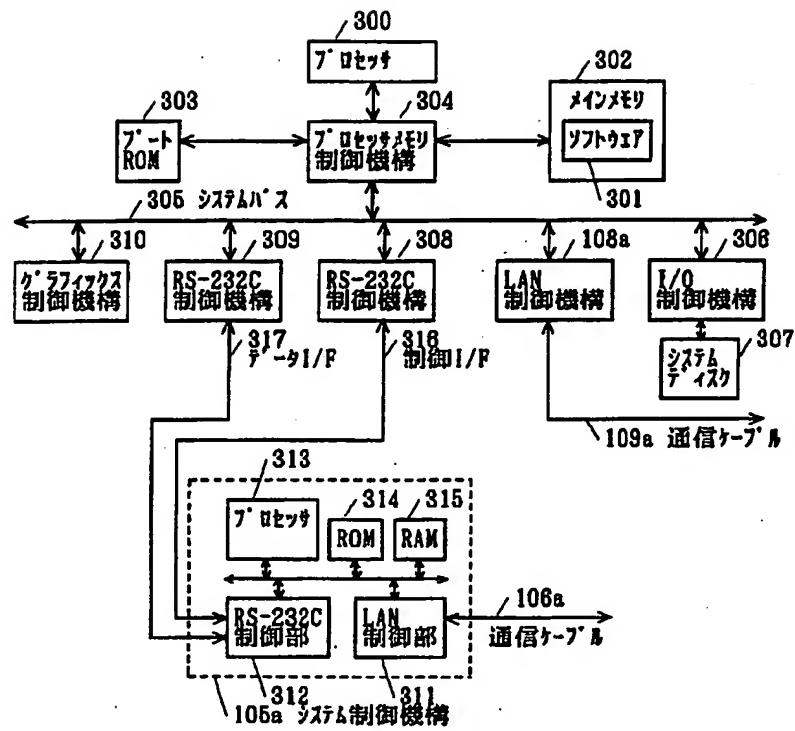
図6





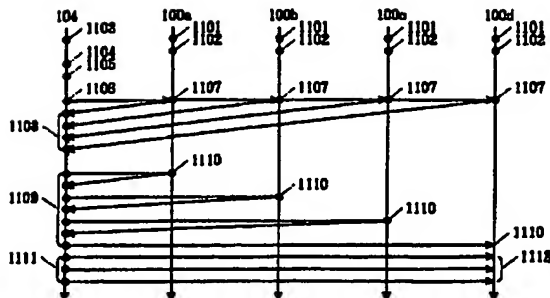
【図3】

図 3



【図11】

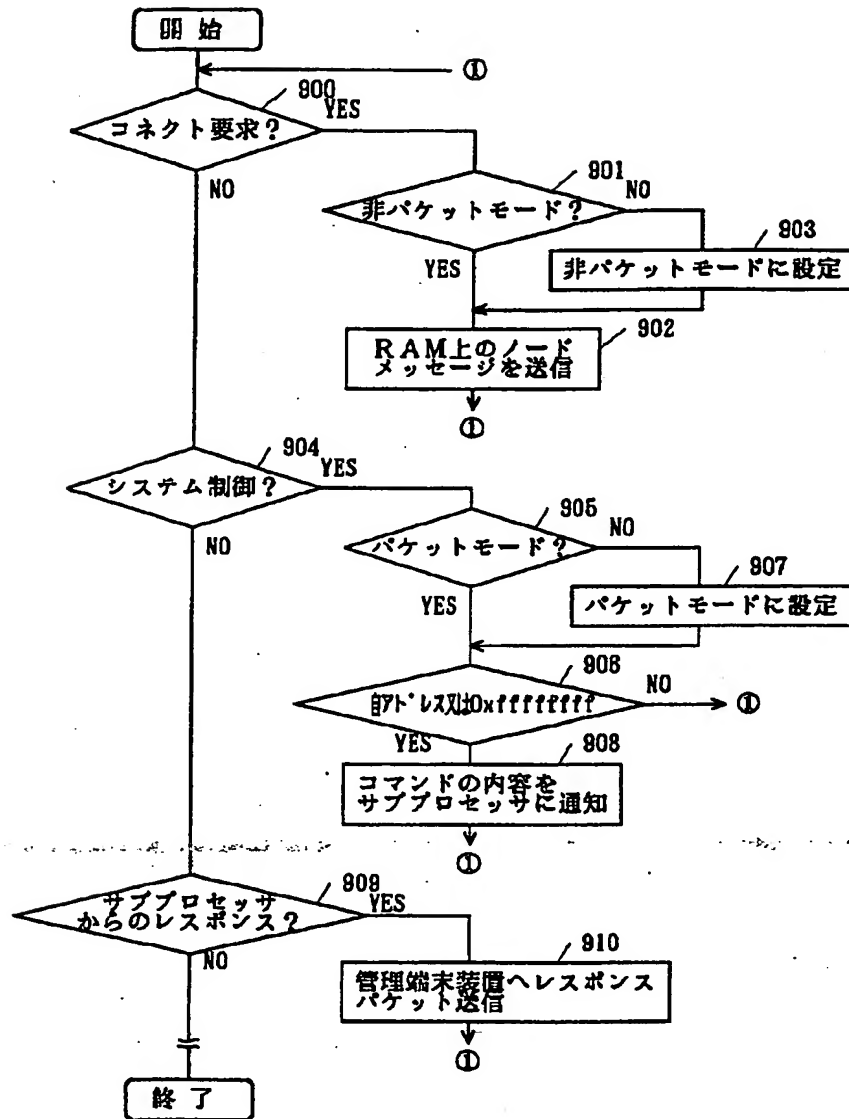
図 11





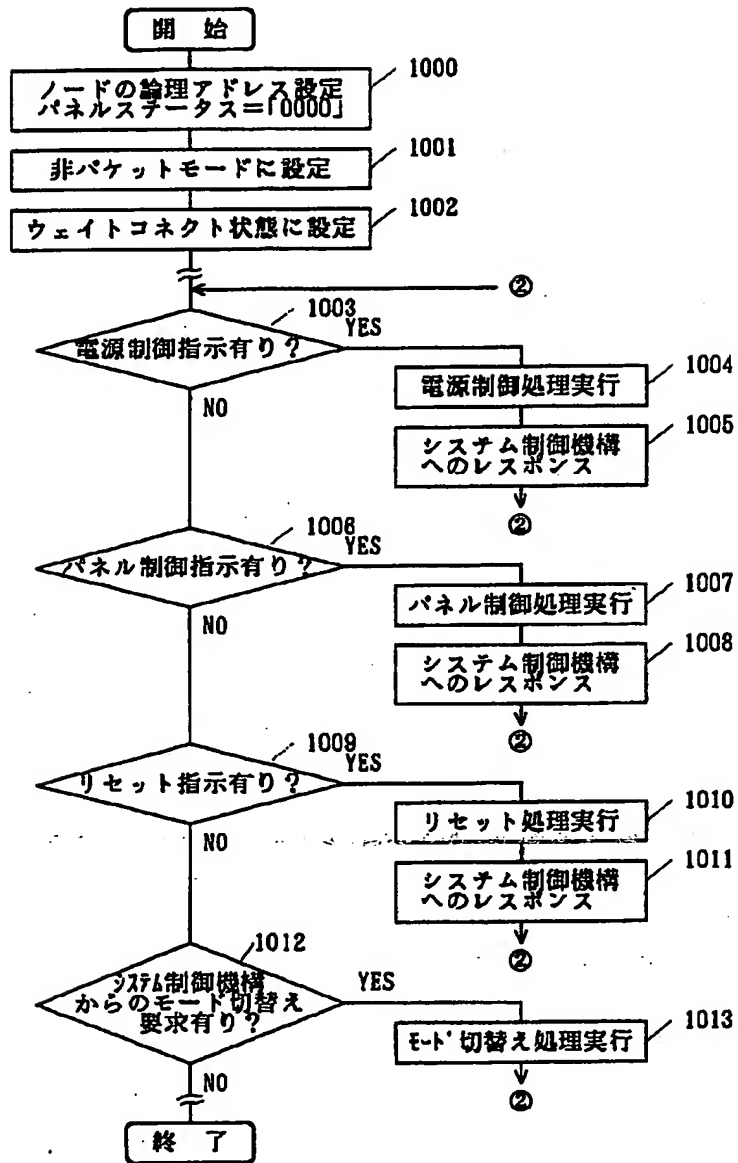
【図9】

図 9



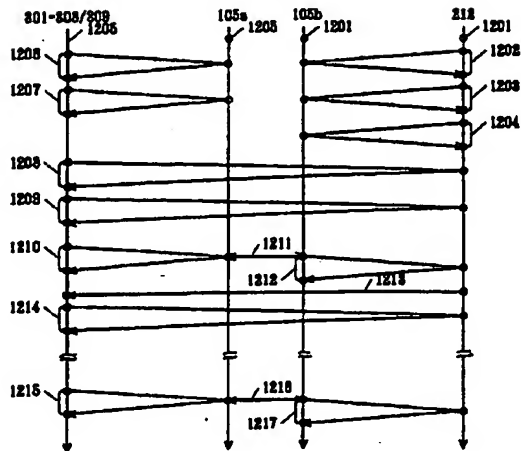
【図10】

図10



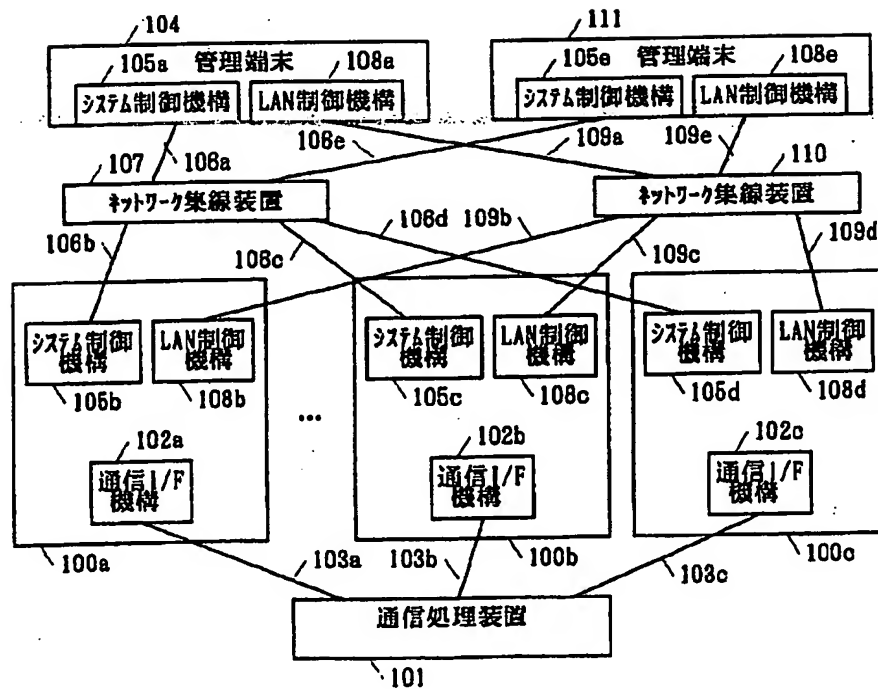
【図12】

図12



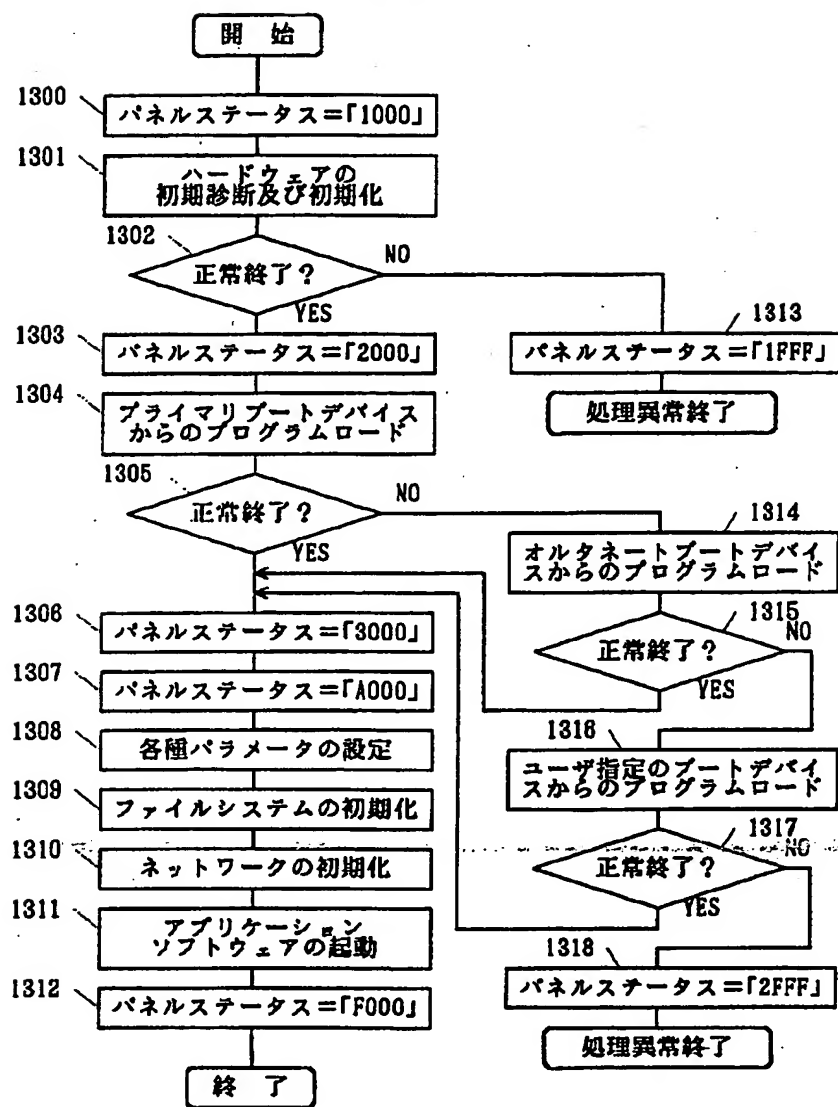
【図15】

図15



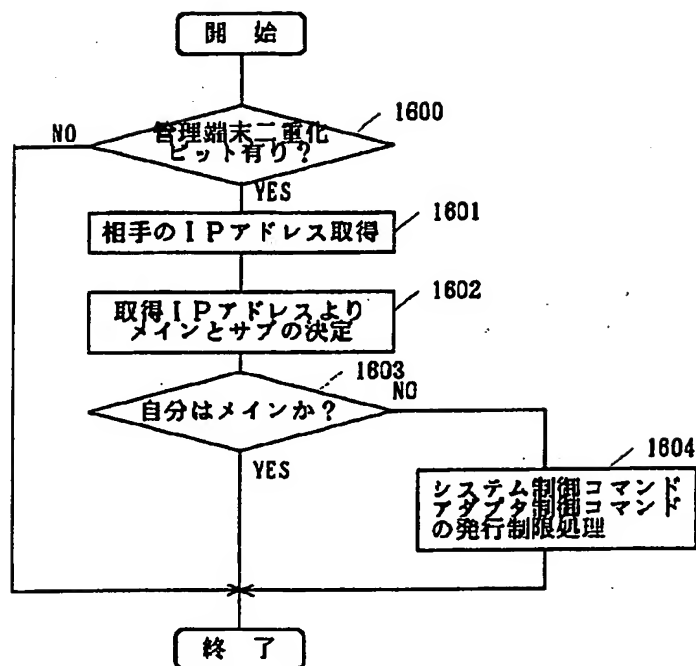
【図13】

図 13



【図16】

図 16





【図17】

図 17

